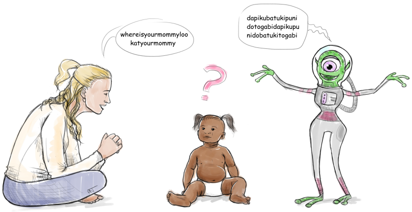
# THE EVOLUTION of LANGUAGE

Proceedings of the Joint Conference on Language Evolution (JCoLE)
Kanazawa, Japan 2022



*Editors*
Andrea Ravignani
Rie Asano
Daria Valente
Francesco Ferretti
Stefan Hartmann
Misato Hayashi
Yannick Jadoul
Mauricio Martins
Yoshei Oseki
Evelina Daniela Rodrigues
Olga Vasileva
Slawomir Wacewicz

EVOLANG
Evolution of Language International Conferences
PROTOLANG
Evolinguistics

# The Evolution of Language

## Proceedings of the Joint Conference on Language Evolution (JCoLE)

Editors:

Andrea Ravignani
Rie Asano
Daria Valente
Francesco Ferretti
Stefan Hartmann
Misato Hayashi
Yannick Jadoul
Mauricio Martins
Yohei Oseki
Evelina Daniela Rodrigues
Olga Vasileva
Sławomir Wacewicz

JCoLE Scientific Committee
September 5, 2022

**Published by:**
Joint Conference on Language Evolution (JCoLE)

Max Planck Institute for Psycholinguistics
Wundtlaan 1
6525 XD Nijmegen
The Netherlands

**Proceedings of the Joint Conference on Language Evolution (JCoLE)**

ISSN 2666-917X
doi:10.17617/2.3398549

# Prefaces

No one thought that September 2022 would be the best time to host an international conference. But we were proud to have hosted it. Remember, when we hosted Evolng in Kyoto, Japan in 2012, it was only one year since the devastating earthquake and nuclear accident in Fukushima. Nevertheless, researchers from all over the world contributed greatly to the progress of the field by engaging in lively discussions about the origin and evolution of language. All participants must have keenly felt that science can only advance when scientists gather face-to-face for discussions.

And this year. For those of us interested in the origin and evolution of language, the last two and a half years have been particularly challenging. Coronaviruses have threatened our way of life and severely interfered with face-to-face communication. This is due to our language that developed technologies for telecommunication and international travels. This is so pity. The war that began in Eastern Europe cruelly shattered the assumption that scientific and technological progress should make us peaceful beings. This is also due to our language. This also is so pity. That is why we decided that this is the year we should hold an international conference face-to-face.

It was during the corona disaster that this plan emerged. But frankly, we did not expect the corona disaster to last this long. Nor did we expect a war to break out in Eastern Europe. And the situation itself was getting more and more difficult. Nevertheless, we chose to hold this international conference.

It is precisely because of this situation that researchers interested in the origin and evolution of language should get together. I wanted to think about what exactly language can do for us in this situation. Does language make us happy or unhappy? What kind of communication is possible for us? Can we have productive discussions with people face to face or electronically? This international conference provided a unique opportunity to test these questions in action. Everyone will have his or her own impressions. What we gained will be different for each of us. Nevertheless, we scientists are happy to have had this conference. We will never give up discussing.

I would like to thank people in Evolang, Protolang, and Evolinguistics for their support of this idea. I thank all my colleagues in Evolinguistics that made this possible. I thank Professor Takashi Hashimoto for making this happen in the beautiful city of Kanazawa. I am particularly thankful to Dr. Rie Asano for her incredible power and skill in getting people together to make this happen. As long as someone like her exists, we can keep science going on.

September 2022
Joint Conference on Language Evolution

*Kazuo Okanoya*, Director

The Joint Conference on Language Evolution was held in Kanazawa, Japan on September 5th – 8th 2022, organised by Evolang, Protolang, and Evolinguistics.

This conference was special in 3 respects. It included in-person attendance, after more than 2 years of online only language evolution events. It pioneered the hybrid modality, with all the benefits - but also logistic issues - it entails. It was the first time in the history of language evolution research that three of the main societies/bodies (Evolang, Protolang and Evolinguistics) came together to organize a joint conference.

As always, the success of the event was, among other things, facilitated by the team effort and synergy of five bodies: the permanent committees, the local organizers, the hybrid conference design and online support team, the scientific committee, and the panel of reviewers. The *permanent committees* of Evolang (headed by Erica Cartmill and Simon Kirby) and Protolang, and Evolinguistics project leaders, have always been keen to provide advice and support when needed. Many members of the permanent committees are our mentors or peers, and they keep making the world of language evolution a better place. The *local organizers* and the hybrid conference design and online support team have worked hard to make the conference possible. Kudos to this cohesive team for turning a ‘Joint Conference on Language Evolution’ idea into reality. The *scientific committee* was in charge of editing and reviewing all contributed abstracts and papers. The JCoLE scientific committee aimed at a strong involvement of early career researchers in the language evolution research community. The committee this year spanned several countries and institutions, featuring members at various stages including PhD, postdoc and early PI. By building on the expertise of existing members and recruiting new ones, we really tried to achieve diversity of scientific backgrounds. The fields and perspective covered by the scientific committee span, among other things: developmental psychology, scientific communication, classical linguistics, speech sciences, field research (both in humans and other species), cognitive neuroscience, gesture, artificial intelligence, computational modelling, genetics, anthropology, acoustics, and music cognition. The conference received many high quality submissions, making the hard work of our *reviewers* especially important. This volume contains many contributions from various disciplines: syntax, semantics, speech sciences, (developmental) psychology, genetics, bioacoustics, anthropology, animal behaviour, neuroscience, and historical linguistics. In selecting contributions we have adopted two main guidelines. First, reviews of submissions were double-blind, avoiding potential biases. Second, contributions were selected not only based on reviewers' scores, but also on the reviewers' confidence in the topic and the dispersion of their scores in terms of statistical measures. The local organising committee, the hybrid conference design and online support team, the reviewers, the scientific committee and the permanent committees all voluntarily contribute their time. Thanks to all for acting in coordination to ensure we did the best possible job we could, considering our other commitments.

Members from Evolang, Protolang and Evolinguistics will surely find novelties and differences in JCoLE 2022 when compared to previous editions of individual conferences. We hope you all enjoy many aspects of those, and let us kindly know about possible improvements.

The programme committee members

*Andrea Ravignani*, *Rie Asano*, *Daria Valente*, *Francesco Ferretti*, *Stefan, Hartmann*, *Misato Hayashi*, *Yannick Jadoul*, *Mauricio Martins*, *Yohei Oseki*, *Evelina Daniela Rodrigues*, *Olga Vasileva*, *Sławomir Wacewicz*

# Permanent Organizing Committees

## Evolang

- Andreas Baumann
- Rudolf Botha
- Christine Cuskley
- Erica Cartmill
- Jean-Louis Dessalles
- Ramon Ferrer i Cancho
- Tecumseh Fitch
- Jim Hurford
- Simon Kirby
- Chris Knight
- Heidi Lyn
- Luke McCrohon
- Kazuo Okanoya
- Thom Scott-Phillips
- Andrea Ravignani
- Nikolaus Ritt
- Kenny Smith
- Maggie Tallerman
- Natalie Uomini
- Sławomir Wacewicz
- Przemysław Żywiczyński

## Protolang

- Ines Adornetti
- Cedric Boeckx
- Piotr Chruszczewski
- Evelina Daniela Rodrigues
- Francesco Ferretti
- Nathalie Gontier
- Stefan Hartmann
- Sverker Johansson
- Luke McCrohon
- Sylwester Orzechowski
- Natalie Uomini
- Olga Vasileva
- Sławomir Wacewicz
- Przemysław Żywiczyński

## Evolinguistics

- Kazuo Okanoya
- Koji Fujita
- Yasuo Ihara
- Harumi Kobayashi
- Takashi Hashimoto
- Kazumi Taniguchi
- Hiroki Koda
- Michio Nakamura
- Yuki Hirose
- Reiji Suzuki

# Conference organization

## Conference director

- Kazuo Okanoya

## Coordinator

- Rie Asano

## Local organizing committee

- Takashi Hashimoto (Chair)
- Koji Fujita (Co-chair)
- Chikako Saeki
- Kazumi Taniguchi
- Yuko Yoshimura
- Tojo Satoshi
- Shohei Hidaka
- Masaharu Mizumoto
- Shun Kurokawa
- Matoba Ryuichi
- Misato Hayashi
- Miwa Shichiri
- Yulri Nonaka

## Hybrid conference design

- Rie Asano
- Tzu-Yin Chen
- Takashi Hashimoto

## Online conference organization team

- Runa Ya (Team leader)
- Elif Yildiz (Team leader)
- Rie Asano
- Tab Bellmann
- Pedro Tiago Martins
- Hironori Nakatani
- Sina Sachse
- Genta Toya

# Panel of Reviewers

Natasha Abner
Ines Adornetti
Daniela Altavilla
Takaya Arita
Mark Aronoff
Rie Asano
Jaume Baixeries
Lluis Barcelo-Coblijn
Andreas Baumann
Ted Bayne
Veronika Beeck
Christina Behme
Teresa Bejarano
Antonio Benítez-Burraco
Aleksandrs Berdicevskis
Balthasar Bickel
Ermenegildo Bidese
Richard Blythe
Cedric Boeckx
Monika Boruta
Hannah Brown
Claudio Brozzoli
Federica Cavicchio
Alexandre Celma-Miralles
Alessandra Chiera
Hélène Cochet
Chris Code
Chris Cummins
Christine Cuskley
Robert Damper
Emily Davis
Bart de Boer
Chiara De Gregorio
Marianne de Heer Kloots
Koen de Reus
Connie de Vos
Dan Dediu
Tania Delgado
Valentina Deriu
Jean-Louis Dessalles
Mark Dingemanse
Ellie Donnelly
Nataliia Drozhashchikh
Shelece Easterday
Axel G. Ekström
Livio Favaro
Nicolas Fay
Andrew Feeney
Francesco Ferretti
Molly Flaherty
Michael Franke
Koji Fujita
Marco Gamba
Mchael Gasser
Chris Golston
Tao Gong
Nathalie Gontier
Michał Głuszkowski
Harald Hammarström
Stefan Hartmann
Takashi Hashimoto
Misato Hayashi
Yuki Hirose
Michio Hosaka
Koji Hoshi
Carla Hudson Kam
Julia Hyland Bruno
Yasuo Ihara
Aritz Irurtzun
Toshiki Iwabuchi
Yannick Jadoul
Gerhard Jaeger
Niklas Johansson
Sverker Johansson
Dariusz Kalociński
Andres Karjus
Noriko Katsu
Henri Kauhanen
Koji Kawahara
Risa Kikuchi
Say Young Kim
Yukinori Kimoto
Simon Kirby
Mirjam Knörnschild
Harumi Kobayashi
Hiroki Koda
Osamu Kondo
Takeshi Konno
Marion Laporte
Kieu Van Le Thi
Shiri Lev-Ari
Simon Levy
Hannah Little
Gary Lupyan
Hannah Lutzenberger
Heidi Lyn
Melanie Malzahn
Mauricio Martins
Pedro Tiago Martins
Hiromi Matsumae
Daiki Matsumoto
Theresa Matzinger
Marco Mazzeo
Alexander Mehler
Gregory Mills
Alexandre Miralles
Steven Moran
Emily Morgan
Olivier Morin
Masahito Morita
Tetsuharu Moriya
Katie Mudd
Elliot Murphy
Thomas Müller
Tomoya Nakai
Michio Nakamura
Hironori Nakatani
Savithry Namboodiripad
Kuniya Nasukawa
Bill Noble
Jonas Nölle
Shinri Ohta
Kazuo Okanoya
Ayaka Onohara
Sylwester Orzechowski
Yohei Oseki
Joel Parthemore
Gerald Penn
Irene Pepperberg
Marcus Perlman
Simone Pika
Marek Placiński
Michael Pleyer
Ljiljana Progovac
Monika Pukli
Sonia Ragir
Andrea Ravignani
Limor Raviv
Gareth Roberts
Sean Roberts
Evelina Daniela Rodrigues
Kazutoshi Sasahara
Sabrina Schalz
Niels Schiller
Robert Scholz
Thom Scott-Phillips
Jose Segovia Martin
Yoshimasa Seki
Arturs Semenuks
Sharita Sharma
Edward Ruoyang Shi
Marta Sibierska
Andrew Smith
Kenny Smith
Göran Sonesson
Rose Stamp
Shane Steinert-Threlkeld
Yushi Sugimoto
Reiji Suzuki
Takuya Takahashi
Miki Takahasi
Jun Takakura
Masanori Takezawa
Monica Tamariz
Shin-ichi Tanaka
Valeria Torti
Genta Toya
Patrick C. Trettenbrein
Ryoko Uno
Martti Vainio
Daria Valente
Freek Van de Velde
Olga Vasileva
Laura Verga
Tessa Verhoef
Marilyn Vihman
Sławomir Wacewicz
Mikihiko Wada
Stuart Watson
Søren Wichmann
James Winters
Marieke Woensdregt
Hiroto Yonenoh
Shoko Yuki
Dietmar Zaefferer
Anna Zanoli
Elizabeth Qing Zhang
Przemysław Żywiczyński

# Contents

# Plenary Presentations

# HUNTER-GATHERERS OF WORDS

Cedric Boeckx*[1,2,3]

*Corresponding Author: cedric.boeckx@ub.edu
[1]ICREA (Catalan Institute for Advanced Studies and Research), Spain
[2]Universitat de Barcelona, Section of General Linguistics, Barcelona, Spain
[3]Universitat de Barcelona Institute for Complex Systems (UBICS), Barcelona, Spain

In this talk I'll be asking what, if anything, the past decade of progress in ancient genomics has been able to tell us about the evolution of human language and cognition. While it has clearly been telling us that a sudden emergence/cognitive revolution scenario (Berwick and Chomsky 2016) is (at best) highly unlikely, it seems to me to also call for a more nuanced scenario than the currently favored alternative ("recognizably modern language is likely an ancient feature of our genus pre-dating at least the common ancestor of modern humans and Neandertals about half a million years ago", as argued in Dediu and Levinson 2013).

The case I will make for this more nuanced narrative builds on an examination of regions of modern human genes associated with signals of positive selection and/or depleted from signatures of introgression from other hominins. Collectively, data from these regions point to neural changes that impacted the way we came to share information and learn from others. If correct, this suggests that the languages/grammars that came to be must have had a distinct profile from previous communicative systems.

## References

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. MIT press.

Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in psychology*, *4*, 397.

# WHAT CAN WE LEARN FROM BIRD SONGS AND RAT TWEETS?

KAZUO OKANOYA*[1,2]

*Corresponding Author: kazuookanoya@gmail.com
[1]Advanced Comprehensive Research Institute, Teikyo University, Tokyo, Japan
[2]Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

## 1. Comparative Approach

Just like our physical body is a product of evolution, so does our behavior. Then it follows that our speech is also a product of evolution. Speech utilizes the respiration system, vocal apparatus, and the brain, all of which are directly descended from our ancestors. Here I introduce two lines of research, one is on birdsong and the other on rat tweets. These studies should provide with pre-adaptations to speech in animal models (Okanoya, 2007).

## 2. Domestication and cultural evolution of complex songs in finches

Juvenile birds learn their courtship songs from their conspecifics through the motor-auditory feedback. Bengalese finches (BFs) are a domesticated strain of wild white-rumped munias (WRMs) imported from China to Japan 270 years ago. BF songs are composed of multiple chunks and each chunk is a combination of 2-4 song notes. Furthermore, chunks are arranged in a finite-state probabilistic automaton. We studied how and why BFs sing such complex songs. We found the following facts. 1) The ancestral strain sings simpler songs. 2) There is high learning specificity in WRMs but not in BFs. 3) BFs have larger song control nuclei and higher level of glutamate receptor gene expressions than WRMs (. 4) Both BF and WRM females prefer complex songs as measured by the nest string assay and males with complex songs are physically fitted than the males with simpler songs. These results promoted sexual selection scenario of song complexity in BFs (Okanoya, 2004).

We further examined factors related with domestication. We examined songs of WRMs in subpopulations of Taiwan. Where there is a sympatric species to

WRMs, songs were simpler. This leads to a hypothesis that in the wild songs needed to be simple to secure species identification, but under domestication this constrains was set free. We also examined socio-emotional indexes including neophobic tendency, tameness, behavioral stress reactions, and corticosterone levels. All indexes suggested that WRMs have higher level of stress and social shyness, which should be adaptive under natural environment, but could be limiting opportunities for learning complex songs (Okanoya, 2017).

To strengthen our hypothesis, we conducted a cultural evolution study. In zebra finches, isolated songs converge into wild type songs after a few generations (Fehr et al, 2009). We isolated juveniles of both strains and reared only by mothers. We used these birds that did not listen to the species-specific songs as founders. After 5 generations, BF songs diverged even more than the founders and never converged. After 3 generations, WRM songs converged into simple patterns, but these were not similar to the wild songs. Results are consistent with the hypothesis on innate constraints of learning in these two strains.

Thus, evolution of song complexity involves not only factors related with strengthen of sexual selection and relaxation of species identification, but also socio- emotional factors due to domestication. Furthermore, recent suggestion of ‘neural crest’ hypothesis that might account for the domestication syndrome fits well with the properties of Bengalese finches. These results on Bengalese finches must be useful in discussing possible biological origin of human speech in terms of proximate and ultimate factors.

## 2. Auditory, cognitive, and brain processing of emotional calls in rats

Rats emit distinct emotional calls under specific situations. When they are in a positive state, such as while copulating or being tickled, they emit 50 kHz pleasure calls. When they are in a distress state, such as when being bitten or received an electric shock, they emit 22 kHz calls. Playing back 50 kHz calls attract rats while playing back 20 kHz calls freeze them.

We studied how these calls were perceived, changed cognitive states, and processed in the brain. We first examined which acoustical parameters were salient in making discrimination of the two categories of the calls. Rats were trained by an operant go-left, go-right discrimination task. They were then tested with various synthetic calls to find out the salient parameters. We found that the frequency (pitch) is the most salient, followed by duration, and then modulation (Saito et al, 2019).

Next, we trained rats to discriminate between two pure tones by the same operant procedure. When they learned the discrimination, they were tested with pure tones of the ambiguous frequencies. Rats responded to the negative lever after exposed to the 22 kHz call while they responded to the positive lever after exposed to the 50 kHz call. Thus, their cognitive states were also influenced by these emotion calls (Saito et al, 2017).

Finally, we tested how these calls were analyzed in the brain. We recorded ERPs from the rat anterior cingulate cortex. ERP amplitudes were stronger to naturally modulated calls than artificial pure tones, and the 50 kHz call emitted stronger response than the 22 kHz call. Taken together, rat emotional calls trigger cognitive bias in the hearer and processed categorically in the brain.

## 2. Putting these together

The birdsong study examined how sequential expertise evolved under sexual selection and domestication. The rat tweet study examined how emotions are conveyed in tweets and affect receivers. These two lines of studies point to important aspects of human speech, namely sequential nature and prosodic modulation.

## Acknowledgements

Work supported by JST, JSPS, and MEXT to KO. I thank all co-workers and students who made these projects possible.

## References

Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. Nature, 459(7246), 564-568.

Okanoya, K. (2004). Song syntax in Bengalese finches: proximate and ultimate analyses. Advances in the Study of Behavior, 34, 297-346.

Okanoya, K. (2007). Language evolution and an emergent property. Current opinion in neurobiology, 17(2), 271-276.

Okanoya, K. (2017). Sexual communication and domestication may give rise to the signal complexity necessary for the emergence of language: An indication from songbird studies. Psychonomic bulletin & review, 24(1), 106-110.

Saito, Y., Yuki, S., Seki, Y., Kagawa, H., & Okanoya, K. (2016). Cognitive bias in rats evoked by ultrasonic vocalizations suggests emotional contagion. Behavioural processes, 132, 5-11.

Saito, Y., Tachibana, R. O., & Okanoya, K. (2019). Acoustical cues for perception of emotional vocalizations in rats. Scientific reports, 9(1), 1-9.

# A WORLD OF SIGN LANGUAGES

CAROL PADDEN

*Department of Communication, University of California San Diego, USA*

The project of sign language research since its start in the middle of the last century has been to document and describe sign languages of different populations around the world. More recently, this body of research has expanded to include sign languages in small community populations: a village, a town, or a family. Even smaller is the sign language of a single deaf signer, called "home sign" where family and sometimes community members participate in sign communication with this individual. Home signers have been reported in nearly every habitable continent in the world. This project has identified as many as 215 different sign languages, large and small, descriptions of which can be found on glottolog.org. The work on smaller sign languages has been especially fruitful in observing emergence of linguistic structures across as brief as one to three generations of signers. Using our work on a small sign language, Al-Sayyid Bedouin Sign Language, my colleagues and I have joined others in describing how linguistic structure unfolds over time (Meir et al., 2017). Faced with this diversity of languages, our field has debated issues of nomenclature for sign languages and how to describe social and cultural factors that play a role in shaping linguistic features of urban and rural or national and village sign languages. Most of these investigations of community properties focus on sign languages where the presence of deaf people – a single signer, a family, a village, or a new school for deaf children – is the impetus leading to the appearance of a sign language. The world language databases, ethnologue.com and glottolog.org, have categorized primary sign languages as those shared in a community of deaf people, distinct from "auxiliary" or "alternative" sign languages, such as those used by Australian Indigenous communities (Ellis et al., 2019) which exist alongside and on occasion, in place of the spoken language. Their existence (and persistence) despite no significant presence of deaf people in the community calls for an explanation. Lauren Reed (2021) has described the case of a home signer in Papua Guinea who enjoys a network of willing hearing communicative partners in and beyond his family. They find that hearing signers are not simply bystanders but are contributing to a shared lexicon used by deaf home signers in different villages who have little contact with one another. Because some have more mobility, hearing signers are transporting signs across geographic space. These sign languages, which they call “network sign languages” (because there is no deaf community as such), exist at the outer bounds of the more deeply researched deaf community sign languages. But they may yet yield unexpected insights into the vast human capacity to build nascent languages when faced with communicative need. If sign languages have helped us to understand properties of spoken languages, small and nascent sign languages – of all types - may be opportunities for us to track the interweaving of social pattern with language use and transmission, as it unfolds in human time.

## References

Ellis, E. M., Green, J., Kral, I., & Reed, L. W. (2019). 'Mara yurriku': Western Desert sign languages. *Australian Aboriginal Studies*, *2*, 89-111.

Meir, I., Aronoff, M., Börstell, C., Hwang, S., Ilkbasaran, D., Kastner, I., Lepic, R., Lifshitz, A., Padden, C. & W. Sandler (2017) The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition, 158*:189-207.

Reed, L. W. (2021). Sign networks: Nucleated network sign languages and rural homesign in Papua New Guinea. *Language in Society*, 1-35.

# INTERACTION IN ANIMAL COMMUNICATION

SIMONE PIKA [*1]

[*]Corresponding Author: simone.pika@uni-osnabrueck.de
[1] Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Language, one of humans' most distinctive traits, still remains a 'mystery' for evolutionary theory. Recently, Levinson (2006) and Levinson and Holler (2014) proposed that it is not language that makes human communication possible, but a special capacity for social interaction. This capacity is a layered assemblage of different social cognitive skills including joint attention, common ground, collaboration and reasoning about communicative intent (Clark 1996). It involves the specific characteristics of face-to-face interaction, frequent employment of mutual gaze, and the exchange of rapid communicative turns.

Surprisingly, this hypothesis has received relatively little research attention in the fields of Animal Behavior and Comparative Psychology (but see Logue and Stivers 2012; Rossano 2013; Genty et al. 2020; Heesen et al. 2022; Pika et al. 2018). The present paper will draw attention to this promising research avenue by (i) providing an overview of the state of the art, and (ii) introducing useful candidates of human social interaction which may be shared across species and taxa.

## References

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Genty, Emilie, Raphaela Heesen, Jean-Pascal Guéry, Federico Rossano, Klaus Zuberbühler, and Adrian Bangerter. 2020. How apes get into and out of joint actions. Shared intentionality as an interactional achievement. *Interaction Studies* 21 (353-386).

Heesen, Raphaela, Marlen Fröhlich, Christine Sievers, Marieke Woensdregt, and Mark Dingemanse. 2022. Coordinating social action: a primer for the cross-species investigation of communicative repair. *Philosophical Transactions of the Royal Society B* 377 (1859):20210110.

Levinson, Stephen C. 2006. On the human “interaction engine”. In *Roots of Human Sociality: Culture, Cognition and Interaction*, edited by N. J. Enfield and S. C. Levinson. Oxford: Berg.

Levinson, Stephen C., and Judith Holler. 2014. The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B Biological Sciences* 369 (20130302):9.

Logue, David M., and Tanya Stivers. 2012. Squawk in interaction: A primer of conversation analysis for students of animal communication. *Behaviour* 149:1283-1298.

Pika, Simone, Ray Wilkinson, Kobin H. Kendrick, and Sonja C. Vernes. 2018. Taking turns: Bridging the gap between human and animal communication. *Proceedings of the Royal Society B Biological Sciences* 285 (20180598):9.

Rossano, Federico. 2013. Sequence organization and timing of bonobo mother-infant interactions. *Interaction Studies* 14 (2):160-189.

# Invited Presentations

# BEYOND COMMON DESCENT: COOPERATIVE BREEDING AND LANGUAGE EVOLUTION

Judith M. Burkart[*1,2]

[*]judith.burkart@aim.uzh.ch
[1]Department of Anthropology, University of Zurich, Switzerland
[2]Center for the Interdisciplinary Study of Language Evolution ISLE, University of Zurich

Great apes are skillful communicators, and enculturation studies have shown that they have an impressive latent potential to acquire human-like communication systems. Nevertheless, only humans have evolved language, and this requires an explanation. Here we argue that a broader phylogenetic perspective that not only focuses on great apes can provide complementary insights by identifying cases of convergent evolution.

We focus on callitrichid monkeys, who like humans are cooperative breeders, i.e. individuals others than the mothers are significantly involved in raising the group's offspring. Group members thus frequently have to closely coordinate their activities and actions, such as handing over infants from one carrier to another, or deciding who engages in infant care or group defense. Callitrichid monkeys are equipped with a set of socio-cognitive skills and abilities that facilitate this coordination and that show striking similarities to elements of the so-called human interaction engine (Levinson & Holler 2014): reliance on mutual gaze as coordination smoother, turn-taking, intentional prosociality, and tuning in with others at the physiological and behavioral level. Thus, the "callitrichid interaction engine" may have catalyzed the emergence of flexible communication in callitrichid monkeys (Burkart et al. 2022), like in humans.

A short review of vocal communication in callitrichids indeed reveals high plasticity which appears rare in non-human primates. They are highly voluble, have large repertoires, and frequently produce combinations of two or more calls. Some calls are functionally referential and callitrichids use the information about identity, group and sex encoded in the calls. They also have some voluntary control over their vocalizations, engage in vocal turn-taking (ontogenetically learned, supported by adult scaffolding), and show some vocal learning even as adults (accommodation). Infants engage in extended bouts of babbling, which predicts the quality of vocal production at the end of infancy but also elicits social interactions from adults, reflecting the special challenges that growing up in such a cooperative breeding system imposes on immatures (Hrdy & Burkart 2020).

It has been argued that to understand the origin of human language, it is crucial to understand the origin of the sociality that enabled its evolution, the human interaction engine. The work on callitrichid monkeys suggests that cooperative breeding contributed to the emergence of several of the elements typically attributed to the human interaction engine, most importantly perhaps high levels of prosociality that become evident in the human eagerness to share information. Since even enculturated, language trained apes rarely use their skills in declarative ways to inform others, and thus clearly lack a strong and intrinsic *Mitteilungsbedürfnis*, this may well have been crucial for language to evolve. Other elements of the human interaction engine may well be unique to humans, or shared with primates generally. Based on currently available primate data, a good working hypothesis may be that our uniquely human language system could emerge because of our double legacy. On the one hand, as great apes, we have inherited from our closest relatives many of the cognitive prerequisites; on the other hand, because our ancestors, unlike all other great apes, started to engage in cooperative breeding and additional forms of intensive cooperation that led to strong interdependence, prosocial motivational dispositions were convergently added and together paved the way for language to emerge.

## References

Burkart, J. M., Adriaense, J., Brügger, R. K., Miss, F. M., Wierucka, K. & van Schaik, C. P. (2022). A convergent interaction engine: Vocal communication among marmoset monkeys. *Phil Trans B, 377,* 20210098.

Hrdy, S. B. & Burkart, J. M. (2020) The emergence of emotionally modern humans: Implications for language and learning. *Phil Trans B, 375,* 20190499.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Phil Trans B, 369*, 20130302.

# LANGUAGE EVOLUTION - UPDATING THE LENS OF PRIMATE COMMUNICATION

ZANNA CLAY*

*Corresponding Author: zanna.e.clay@durham.ac.uk
[1] Department of Psychology, Durham University, Durham, UK

As our closest living relatives, non-human primates (henceforth primates) provide a unique lens through which we can explore the evolutionary basis for our behavioural and cognitive capacities, including for language. Over the past four decades, there has been intense amount of interest in the communicative capacities of primates and in particular, what they tell us about the evolutionary origins of language. Inspired by seminal work of Seyfarth, Cheney and colleagues (e.g. Seyfarth et al., 1980), the initial focus of interest was mostly on primate vocalisations, particularly monkey alarm calls, and what they could inform about the evolution of semantics, syntax and relatedly social cognition (e.g. Zuberbuhler, 2005). A large body of research has, in particular, been dedicated to the study of *functional reference*, considered an evolutionary precursor to linguistic reference (Clay et al, 2010; Townsend & Manser, 2013). Referential signals must fulfill both the production criteria of being tightly associated with their eliciting stimulus as well as the perception criteria of eliciting the same response in a receiver as the putative referent would do (in its absence) (Macedonia & Evans, 2003). While evidence of functionally referential signals has provided important insights into the evolutionary building blocks of semantics and other linguistic properties, empirical focus on such signals, which by definition are functionally fixed, may have led to a biased view of primate communication which fails to take into account its apparent flexibility (Taylor et al., 2022) and multimodal nature (Liebal et al., 2014). In addition, by focusing so heavily on referential primate calls, the comparative literature has dedicated relatively scant attention towards the affective nature of primate signals, which have long been deemed out of the realms of scientific investigation and even scientifically uninteresting.

Nevertheless, comparative research is now starting to address these imbalances, and as a consequence, offers new and exciting insights into primate communication that enable a richer picture of the complex mosaic of traits which underlie language evolution. In this presentation, I will discuss three key empirical shifts that are starting to occur in the primate communication literature and their relevance for understanding language evolution.

The first is a growing appreciation of the functionally flexible nature of primate vocalisations. While certain linguistic signals are tightly associated with their referents, many are actually highly context-dependent and loosely associated with their referents, enabling them to fulfill a range of functions. We have argued that the same may also be true for primate signals and that very notion of functional reference needs revision in order to account for the important role of context (Scarantino & Clay, 2015) as well as the functionally flexible nature of most primate vocalisations (Taylor et al., 2022).

The second is a greater focus on the multi-modal nature of primate communication and its relevance to language. Until recently, most primate communication research has focused on unimodal signals with often a strong discontinuity in both theoretical and methodological approaches between modalities, particularly vocalisations and gestures (Liebal et al., 2014). Focusing on one modality alone can result in biased views about the evolutionary basis of language. For instance, many theories have focused on a unimodal origin of language, despite the fact that both primate communication and human language are multi-modal. Thus, although such work on multi-modality is still in its infancy, addressing this imbalance in primate research is starting to enable more accurate and balanced models of language origins (Slocombe et al., 2011; Liebal et al., 2014, 2022).

The third shift is a steadily growing interest in the affective basis of primate communication and its relation to intentionality (Heesen et al., 2022). Although intentional communication about affective states is a central part of human communication, there has long been a perceived discontinuity between intentional versus affective signaling in primate communication. In particular, primate vocalisations and facial expressions have been traditionally viewed as involuntary expressions of internal states, whereas gestures are by definition intentional. Nevertheless, such a view is increasingly contentious given evidence of intentional vocal and facial expression production, as well as the apparent void of arousal-based explanations in gesture research. By showing that primate signals can be both affective and intentional, regardless of signal modality, a more dimensional approach can contribute to a richer picture of primate communication and how its affective and cognitive processes have jointly shaped the evolution of language.

## References

Heesen R., Sievers C, Gruber T., Clay, Z. (2022). Primate communication: Affective, intentional, or both? In B. L. Schwartz & M. J. Beran (Eds.), Primate Cognitive Studies. Cambridge University Press.

Liebal, K., Waller, B. M., Slocombe, K. E., & Burrows, A. M. (2014). *Primate communication: a multimodal approach*. Cambridge University Presss

Liebal, K., Slocombe, K. E., & Waller, B. M. (2022). The language void 10 years on: multimodal primate communication research is still uncommon. *Ethology Ecology & Evolution*, *34*(3), 274-287.

Macedonia, J. M., & Evans, C. S. (1993). Essay on contemporary issues in ethology: Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology*, *93*(3), 177-197.

Taylor, D., Clay, Z., Dahl, C. D., Zuberbühler, K., Davila-Ross, M., & Dezecache, G. (2022). Vocal functional flexibility: what it is and why it matters. *Animal Behaviour*, *186*, 93-100.

Townsend, S. W., & Manser, M. B. (2013). Functionally referential communication in mammals: the past, present and the future. *Ethology*, *119*(1), 1-11.

Scarantino, A., & Clay, Z. (2015). Contextually variable signals can be functionally referential. *Animal Behaviour*, *100*(100), e1-e8.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, *210*(4471), 801-803.

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: the need for multimodality in primate communication research. *Animal Behaviour*, *81*(5), 919-924.

Zuberbühler, K. (2005). The phylogenetic roots of language: evidence from primate communication and cognition. *Current Directions in Psychological Science*, *14*(3), 126-130.

# MULTIPLE ATTENTION UNDERPINS THE CO-EVOLUTION OF THOUGHT AND COMMUNICATION

KOJI FUJITA

fujita.koji.3x@kyoto-u.ac.jp
Graduate School of Human and Environmental Studies,
Kyoto University, Kyoto, Japan

Human language sharply differs from animal communication systems in that it is adaptive not only for communication but also for creative, complex thinking (internal mental calculations such as inference, planning, imagination, etc.), but it is a totally different issue, still remaining open, whether language initially evolved as a tool of communication or as a tool of thought. While many researchers, linguists and biologists alike, opt for the communication view, the opposite view, that language was originally an internal tool of private thought and only later reused for communicative purposes when coupled with some means of externalizing such thoughts (gestures, signs, vocalizations, etc.), is also favored by linguists of a certain theoretical persuasion (Berwick & Chomsky, 2016). Both views have their own merits and demerits, and it is worth considering the possibility of unifying them into an integrative co-evolutionary theory of thought and thought. I will argue that "multiple attention," here defined as a cognitive capacity for paying and maintaining attention to more than one object, provides a key to such an integration.

It is important to note that language is a complex trait consisting of several subfunctions each of which may have evolved in other species independently before they were linked together to form the human language faculty. In other words, there is no single precursor to human language, hence no single original function. Different views stem from which of these subfunctions one focuses on. While gestures and vocalization, as well as cooperativeness, prosociality and mind reading, are obviously linked to communicative functions, the kind of hierarchical syntax that is sometimes taken to be the hallmark of human language (structure dependence) may not be so. Rather, its contribution to complex, hierarchical thought seems more likely, given in particular that structure dependence often gives rise to structural ambiguity such that the same linear strings of words have more than one semantic interpretation depending on which hierarchical structure they derive from, which hinders efficient communication. Thus there is an apparent tension or dilemma between complex thought and communication.

A close examination of linguistic hierarchical structure, in particular of how complex structures are dynamically built up from scratch, reveals that there is a common cognitive underpinning for structure dependence and social communicative behavior; namely, one needs to pay attention to multiple objects as a target of internal mental manipulation, freely switching between them. This is illustrated by comparing a basic linguistic structure like [[ the dog ][ saw [ the cat ]]] (where *the dog* and *the cat* are derived separately, to be later integrated into a single structure) and intention sharing (or shared intentionality; Tomasello, 2008) between a speaker and a hearer (each takes note of the other's intent and tries to conform). I suggest that this kind of multiple attention allowed the co-evolution of complex thought (by hierarchical linguistic structure) and communication (by mind-reading for intention sharing).

Previously, the thought-first view was advanced on certain debatable assumptions about language evolution, in particular that there was no protolanguage, and that human language came into being abruptly when the capacity for hierarchical structure building (sometimes known as "Merge") emerged by some mutation, without any precursor. There are other more natural assumptions we should take into serious account which reconcile the thought view with the communication view. It can be argued, for instance, that there was a protolanguage stage where it already served for the purpose of both thought and communication, and that the shift from linear protolanguage (Jackendoff & Wittenberg, 2016) to hierarchical human language changed human thought more drastically than human communication. More pertinent to the present discussion, however, is the idea that hierarchical syntax is supported by the same cognitive trait that underlies human communication.

In such a co-evolutionary scenario, the role of externalization cannot be overemphasized. Rather than being a subsidiary process added later, it was the core feature that made it possible to manipulate abstract linguistic concepts on a par with concrete objects to create more and more complex thoughts. Language is more than communication, but it could not have evolved without communication via externalization.

## References

Berwick, R.C., & Chomsky, N. (2016). *Why only us: Language and evolution.* Cambridge, MA: MIT Press.

Fujita, H., & Fujita, K. (2021). Human language evolution: a view from theoretical linguistics on how syntax and the lexicon first came into being. *Primates,* online. DOI:10.1007/s10329-021-00891-0

Fujita, K. (2017). On the parallel evolution of syntax and lexicon: a Merge-only view. *Journal of Neurolinguistics, 43,* 178-192.

Greenfield, P.M. (1991). Language, tools, and brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences, 14,* 531-595.

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298,* 1569-1578.

Hauser, M.D., & Watumull, J. (2017). The universal generative faculty: the source of our expressive power in language, mathematics, morality, and music. *Journal of Neurolinguistics, 43,* 78-94.

Jackendoff, R., & Wittenberg, E. (2016). Linear grammar as a possible stepping-stone in the evolution of language. *Psychonomic Bulletin & Review, 24,* 219-224.

Martins, P.T., & Boeckx, C. (2019). Language evolution and complexity considerations: the no half-Merge fallacy. *PLoS Biology, 17(11),* e3000389.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT Press.

# EMERGENT CONSTRUCTIVE APPROACH TO EVOLINGUISTICS

TAKASHI HASHIMOTO

hash@jaist.ac.jp
Japan Advanced Institute of Science and Technology, Japan

Evolinguitics aims to deepen our understanding of humans from an evolutionary perspective by clarifying the origins and evolution of language and communication, emphasizing two idiosyncratic natures of human linguistic communication. They are to use hierarchically organized symbol sequences in language and to share intentions in communication. We believe that integrating these two characteristics makes humans co-creative and intelligent. In this talk, I introduce two studies taking an emergent complex approach. It is a methodology to analyze complex systems by constructing and operating the evolutionary and emergent process of complex phenomena and plays a vital role in the attempt of Evolinguistics. One research I introduce is an evolutionary simulation of recursive combination, which is thought of as the essential ability to form hierarchical structures in language and many human cultural products. The other is an experimental semiotic study to consider the process, mechanisms, and neural basis of symbolic communication systems. Finally, I propose a hypothesis that understanding others' intentions is realized by abduction, which consists of hypothesis generation and their selection, and that the former is brought by recursive combination and the latter by embodiment. This hypothesis is a sort of integration of hierarchy and intention sharing.

# HISTORIAL AND ECOLOGICAL ASPECTS OF THE EVOLUTION OF HUMAN LANGUAGE FACULTY

YASUO IHARA

iharay@bs.s.u-tokyo.ac.jp
Department of Biological Sciences, University of Tokyo, Tokyo, Japan

Any meaningful hypothesis to account for the emergence of language faculty needs to be congruent with our knowledge about the history of human evolution, which is continuously updated based on paleoanthropological researches. It also should consider the ecology of human ancestors, given that natural selection has shaped functional traits as adaptation to their physical as well as social environment. During the last five years, I have participated in the Evolinguistics Project as a member of the Human Evolution Team. The key role that our team has played in the project has been to bring historical and ecological perspectives to the study of language evolution. I will talk about some of the issues that we have worked on in this research program. As a basic premise, I consider the evolution of language faculty as a long-term, step-by-step process through the seven-million-year history of human evolution, rather than a single major event. I also assume that sub-functions composing the language faculty emerged sequentially during the process, some of which may have emerged, perhaps long before language itself, as preadaptations.

For the purpose of summarizing the relevant historical framework as simply as possible, without being overwhelmed by the vast detailed information, I propose a highly simplified three-phase scheme for the history of human evolution (Ihara, forthcoming). Phase 1 of the scheme corresponds to the first several million years since the human-chimpanzee divergence. Among the African apes at that time, the early hominins were unique in that they were bipedal and had reduced canine teeth. The transition to Phase 2 is marked by the beginning of the oldest stone-tool industry. During Phase 2 there was a dramatic increase of brain size, which chronologically parallels increasing sophistication of tool-making technologies. This is also when early *Homo* diffused out of

Africa for the first time in the human history. A second phase transition is equated with the advent of novel types of artifacts, such as personal ornaments and arts, so-called modern human behavior. For example, one of the oldest signs of modern human behavior is the 75,000-year-old shell beads from South Africa.

I further argue that each of these phases may be characterized by a distinct dominant evolutionary force, and that they were not just temporally consecutive, but also causally connected (Ihara, forthcoming). To be specific, I speculate that social and life-history evolution that occurred in Phase 1 set the stage for cognitive evolution in Phase 2, which enable cultural evolution in Phase 3. I am aware that this is all too simplistic, and I do not intend to suggest, for example, that social and life-history evolution took place only in Phase 1. Nevertheless, this kind of schematization may hopefully help to draw attention to the possible role of niche construction in human evolution and a sort of evolutionary cascade that may have resulted to drive phenotypic diversification between modern humans and non-human apes.

I will discuss some specific research topics that have been studied in the Human Evolution Team, including hominin confrontational scavenging (Nakamura et al., 2019), manufacture of composite tools (Sano et al., 2019), social selection to favor reduced aggression (Ihara, 2020), and hierarchical object manipulation (Hayashi & Takeshita, 2022).

## Acknowledgement

This work was supported by KAKENHI JP17H06381.

## References

Hayashi, M., & Takeshita, H. (2022). Hierarchical object combination and tool use in the great apes and human children. *Primates*.

Ihara, Y. (2020). A mathematical model of social selection favoring reduced aggression. *Behavioral Ecology and Sociobiology 74*, 91.

Ihara, Y. (forthcoming). Early adaptations to human language: Historical and ecological aspects. In Y. Ihara et al. (Eds.), *Human Evolution and Language: A View from Hierarchy and Intention Sharing*. Springer.

Nakamura M. et al. (2019). Wild chimpanzees deprived a leopard of its kill: Implications for the origin of hominin confrontational scavenging. *Journal of Human Evolution 131*, 129-138.

Sano K. et al. (2019). The earliest evidence for mechanically delivered projectile weapons in Europe. *Nature Ecology & Evolution 3*, 1409-1414.

# INFERENTIAL NATURE OF GESTURE AND ITS ROLE IN EVOLUTION OF LANGUAGE

HARUMI KOBAYASHI [*1]

[*]Corresponding Author: h-koba@mail.dendai.ac.jp
[1] Tokyo Denki University, Japan

One important characteristic of human gesture is its inferential nature, that is, the meaning of gesture must be inferred by the recipient. This inferential nature can be observed in almost every aspect of human communication, therefore, examining this inferential nature of gesture must contribute to the exploration of the origins and the evolution of language. Based on experimental evidence from our laboratory observations, I discuss the inferential nature of pointing gestures. To interpret pointing gesture, it is necessary to recognize the accompanying characteristics of pointing gestures and make relevant inferences in the given context. The present results show that both children and adults are sensitive to inferential nature of pointing gesture and appropriately use the information for word learning. These results suggest that inferential nature of human communication played an important role in evolution of language.

# VOCAL LEARNING AND THE EVOLUTION OF BEAT-BASED RHYTHMIC PROCESSING

ANIRUDDH D. PATEL*[1]

*Corresponding Author: a.patel@tufts.edu
[1]Dept. of Psychology, Tufts University, Medford, MA, USA

Rhythms with a periodic beat are widespread in human music and are not part of ordinary language. Yet beat-based rhythmic processing may have deep evolutionary and neurological links to spoken language. About 15 years ago I hypothesized that human beat processing built on the neural circuitry for complex vocal learning, an ability foundational to the evolution of speech (Patel, 2006). In this talk I review and update this hypothesis in light of cross-species, neural, and genetic research since that time, including research on nonhuman primates, birds, pinnipeds, humans, and rodents (Patel, 2021). Based on this research I maintain that there are important evolutionary and neural relations between vocal learning and beat-based rhythmic processing. I also hypothesize that gene-culture coevolution elaborated on these links to produce neural specializations for beat processing in the dorsal auditory stream of the human brain. I suggest that beat-based rhythmic processing provides a model system for studying cognitive gene-culture coevolution, in which neuroscientific, cross-species, and genetic research can be meaningfully integrated (Patel, in press).

## References

Patel, A.D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, 24:99-104.

Patel, A.D. (2021). Vocal learning as a preadaptation for the evolution of musical beat perception and synchronization. *Philosophical Transactions of the Royal Society B.* 376 (1835), 20200326

Patel, A.D. (in press). Musicality and gene-culture coevolution: Ten concepts to guide productive exploration. In: E.H. Margulis, P. Loui, & D. Loughridge

(Eds.) *The Science Music Borderlands: Reckoning with the Past and Imagining the Future*. MIT Press. https://psyarxiv.com/qp6jx/

# LINKING LANGUAGE EVOLUTION, LANGUAGE ACQUISITION, AND LANGUAGE DIVERSITY: HOW SOCIAL AND COGNITIVE PRESSURES SHAPE LEARNING AND COMMUNICATION

LIMOR RAVIV *1,2

*Corresponding Author: limor.raviv@mail.huji.ac.il

1 Language Evolution and Adaptation in Diverse Situations (LEADS) group, Language & Genetics Department, Max Planck Institute for Psycholinguistics, The Netherlands

2 Centre for Social, Cognitive and Affective Neuroscience (cSCAN), University of Glasgow, UK

What are the social, environmental, and cognitive pressures that shape the evolution of language in our species? Why are there so many different languages in the world? And how did this astonishing linguistic diversity come about? These are some of the most interesting questions in the fields of cognitive science and linguistics, and represent the range of topics discussed in my research so far.

My work focuses on linking core aspects of language acquisition, language evolution, and language diversity using a range of novel behavioral paradigms and computational models. My goals are to (1) shed light on the communicative pressures and cognitive constraints (e.g., memory limitations, input variability, generalization, multimodality) that shape language use and social interaction in our species, and (2) to identify the social, environmental, and cross-cultural factors (e.g., population size, network structure, gender/age distribution, noisy physical conditions) that lead to language diversity and to cross-linguistic variation.

In this talk, I will provide an overview of my research (including methods and results from selected projects), as well as present future directions and ongoing work. Specifically, I will outline previous studies that used group communication experiments and artificial language learning experiments to test the live formation and acquisition of new languages in the lab. I will focus on the effect of population size, namely, the differences between languages that evolved in big vs. small mini-communities in terms of their compositional structure, their rate of change, and their ease of learning. I will then discuss the role of input variability in underlying group size effects and shaping this evolutionary process, as well as shaping humans' learning and generalization more broadly. Finally, I will link these results to the human self-domestication hypothesis for language evolution, highlighting recent work that tests key components of self-domestication (e.g., prosociality, interaction with strangers) in new animal models and in AIs.

# Refereed Contributions

# QUALITY NOT QUANTITY IN SYNONYMY AVOIDANCE

Aja Altenhof[1,2] and Gareth Roberts[*2]

[*]Corresponding Author: gareth.roberts@ling.upenn.edu
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, Scotland
[2]Department of Linguistics, University of Pennsylvania, Philadelphia, USA

Words often compete for meaning and synonymy is extremely common. But *absolute* synonymy is exceedingly rare (Murphy, 2003). How does this situation arise? One possibility is that it results from a cognitive bias against synonymy, such as the mutual exclusivity bias (Markman & Wachtel, 1988; Hurford, 2003). Another possibility is that it is caused by differences in the perceived contextual distribution of candidate synonyms, which become amplified in learning (Hudson Kam & Newport, 2009).

To investigate this we conducted an experiment in which 362 participants were asked to learn three new slang words—presented in example sentences—and (as a cover task) to guess what the words meant. Two of the words (*snater* and *fincur*) were verbs, and participants were told they referred to the same action. The third word (*murp*) was a noun and functioned purely as a distractor. In the initial *Exposure phase*, 36 sentences were presented on screen, one by one, with 10 s between each sentence. Sentences were presented in a random order, with 12 per word. Verbs could appear in negative contexts (e.g., “Don’t fincur in front of me!”), positive contexts (e.g., “My friends and family love my snatering skills.”), or neutral contexts (e.g., “Dogs can’t snater”). The noun *murp* appeared only in neutral contexts.[1]

After the Exposure phase came the *Generalization phase* in which participants were presented with 36 unseen sentences, each with a gap for participants to insert one of the three words. Of these sentences, 24 required verbs (12 positive and 12 negative) and the remaining 12 required nouns.

We manipulated the distribution of contexts in the Exposure phase. In the Neutral condition, all sentences were neutral. In the Random condition, both verbs were evenly distributed across positive and negative contexts. In the Consistent condition, one verb appeared only in positive sentences and the other appeared only in negative sentences. In the Overlapping condition, one verb occurred 75%

---

[1] Sentences were chosen following extensive piloting to ensure that the intended difference in valence existed.

of the time in positive sentences and 25% in negative sentences; the reverse was true for the other verb. For the 75%-positive condition, both verbs occurred 75% of the time in positive sentences and 25% in negative sentences; this distribution was reversed for the 75%-negative condition.

To identify how much participants differentiated the verbs in the Generalization phase we calculated, for each verb, a *word context score* by dividing its frequency in its higher-frequency context by the sum of its frequencies for both contexts. We then multiplied the context scores for the two verbs to get a *Differentiation score*. A Differentiation score of 0 meant that one verb had not been used at all. A score of 1 meant that the two verbs were 100% differentiated. 0.25 meant that both verbs were used equally often in both contexts (equivalent to random assignment). A score around 0.667 indicated "partial differentiation", where one verb was entirely specialized to one context and the other played a more default role and occurred in both (as with, e.g., English *thin : skinny*).

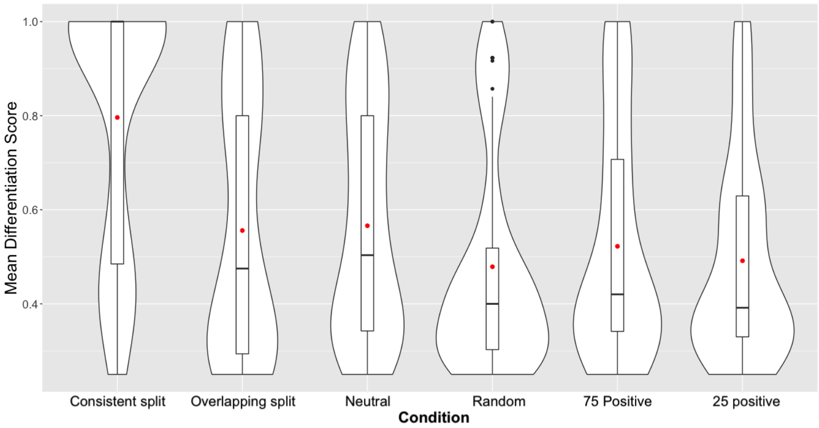


Figure 1. Differentiation score distributions for all conditions. Red dots indicate mean values.

Results are displayed in Figure 1. In the Consistent condition, full (or close-to-full) differentiation occurred. In all other conditions, regardless of Exposure-phase distribution, we found partial differentiation. A replication in which neutral sentences were included in the Generalization phase found the same result.

We consider these results to be consistent with an account based on a cognitive bias against synonymy (cf. Hurford, 2003). Participants differentiated potential synonyms at a greater than chance level and to an extent inconsistent with their distribution. This is not to say that distribution was irrelevant; there was significantly greater differentiation in the Consistent condition. However, it seems that the quality of differentiation in Exposure was more important than quantity.

## References

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66.

Hurford, J. R. (2003). Why synonymy is rare: Fitness is in the speaker. In W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, & J. T. Kim (Eds.), *European conference on artificial life* (pp. 442–451). Dortmund.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms.* Cambridge: Cambridge University Press.

# R YOU SURE THAT YOUR /R/ IS TRILLED? A METHODOLOGICAL CAVEAT

RÉMI ANSELME*[1], FRANÇOIS PELLEGRINO[1], and DAN DEDIU[2]

*Corresponding Author: remi.anselme@univ-lyon2.fr
[1]Laboratoire Dynamique du Langage, CNRS/University of Lyon, Lyon, France
[2]University of Barcelona & Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain

Databases are great tools for typological studies and for understanding the evolution of language, but they can rarely be taken at face value. Many (if not all) databases have various limitations and usage constraints (see, for example, for WALS (Cysouw, Dediu, & Moran, 2012; Hunley, Bowern, & Healy, 2012) and UPSID (Simpson, 1999)), which can "trickle down" to (or even be amplified by) the studies based on them. We focus here on *phonemic databases*, which can be used, among others, to compare phonemic inventories across languages and to make claims about language evolution through perceptual or articulatory mechanisms for instance. However, such databases often suffer from an intrinsic ambiguity between *phoneme/allophone/phone*, which often percolates down to research that aims to study the *phonetic systems* of languages from (supposedly) *phoneme inventories* (for a recent example, see (Winter, Sóskuthy, Perlman, & Dingemanse, 2022)). One common hurdle is that sometimes the phoneme is seen as a monolithic object without variation in its phonetic realizations, and without contextual dependencies. Moreover, *written symbols* can be misleading because the same symbol may represent different phonetic realities (Anderson et al., 2018). To drive these points home, we will focus here on *r* (Barry, 1997).

Not all segments are directly comparable, and the *rhotics* class is a perfect example because of their great phonetic variability (Chabot, 2019; Lindau, 1985). In particular, by default the rhotic *r* is often considered as an alveolar trill: from a *phonemic* point of view, UPSID (Maddieson, 1984) reports that among the languages that have a rhotic, at least half have one alveolar trill or more. But to say that a language has an alveolar trill *phoneme* does not inform about its *potential phonetic realizations*. Several studies show that the trill alveolar phonemes have few of their allophones that are actually trilled (Sebregts, 2014; Blecua, 2002; Rafat, 2010), potentially leading to a distorted picture. As an example, we consider here a recent article (Winter et al., 2022) which correlates the presence of the trilled /r/ sound with "roughness" cross-modally. Their result is supported by several approaches, one being a cross-linguistic statistical study of 332 languages, where, in order to determine whether a language has a trilled /r/ or not, the authors rely on (a) their phonetic judgments based on the literature, and (b) PHOIBLE (Moran & McCloy, 2019). Although their phonetic judgments took precedence over PHOIBLE, both methods tend to force the interpretation of the phoneme as having a single realization and may introduce some residual subjectivity.

As a sort of proof of concept, we decided to reproduce their process, by focusing on a random sub-sample of the languages they considered (34 of 332; $\approx$10%) for which they mention a trilled /r/. Our rerun is based on the authors' own comments in the data files used for data preparation, which we used as a starting point for searching grammars and reports on the languages of interest. When the authors based their judgment on PHOIBLE, we accessed the primary sources in there to assess whether /r/ was indeed trilled (however, this is still problematic because of the limitations mentioned in the previous paragraph). We systematically tried to use several resources and avoid making decisions based on only one source.

The findings on this sub-sample are that, first, in 16 of the languages (47% of the sub-sample), we fail to reach the same conclusions as the authors. There are three languages that should not have been included in the original analysis as they show a contrast between a trilled /r/ and a non-trilled /r/ (exclusion criterion in the original study). Eight other languages should not have been considered as having a trilled /r/ because the rhotic is described as not trilled in the primary sources even if the symbol *r* is used in the grammars (leading to confusion). And in another six languages, the articulation of the rhotic is not defined in the grammars, making it is impossible to tell if it is a trilled /r/ (*sonorant*, *resonant*, *liquid*, or *rhotic*) or not, other than by implicitly assuming an interpretation of the symbol *r* as a trilled /r/ (which is not warranted). For the remaining 18 languages (53% of the sub-sample), we may consider them as having a trilled /r/, since we did find grammars containing phonetic information. Still, this is not always the case, with some decisions made solely on the basis of the mention ”trill” in the phoneme inventory of the languages' reference grammars.

Should our preliminary results generalize to the whole sample of 332 languages, it may question the validity of the statistical cross-linguistic findings of a cross-modal association between the alveolar trill and “roughness”. However, our point here is not to criticize this particular study, but to highlight the dangers of assuming phonetic meaning from written symbols allegedly representing phonological systems. In fact, we want to underline that, while large databases (and their associated statistical methods) are essential for the modern language sciences and the study of language change and evolution, they cannot be used without considerable care and expertise. The good news is that such expertise is available and achievable, as well as incorporable in large-scale databases (see, for example, the inclusion of multiple inventories and of allophones in PHOIBLE), and that new statistical approaches, capable of dealing with the remaining uncertainty, can be developed (probably in a Bayesian framework).

## References

Anderson, C., Tresoldi, T., Chacon, T., Fehn, A.-M., Walworth, M., Forkel, R., & List, J.-M. (2018). A cross-linguistic database of phonetic transcription systems. *Yearbook of the Poznan Linguistic Meeting*, *4*(1), 21–53.

Barry, W. J. (1997). Another r-tickle. *Journal of the International Phonetic Association*, *27*(1-2), 35–45.

Blecua, B. (2002). *Las vibrantes del español: manifestaciones acústicas y procesos fonéticos.* Unpublished doctoral dissertation, Universitat Autònoma de Barcelona.

Chabot, A. (2019). What's wrong with being a rhotic? *Glossa: a journal of general linguistics*, *4*(1), 38.

Cysouw, M., Dediu, D., & Moran, S. (2012). Comment on "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa". *Science*, *335*(6069), 657–657.

Hunley, K., Bowern, C., & Healy, M. (2012). Rejection of a serial founder effects model of genetic and linguistic coevolution. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1736), 2281–2288.

Lindau, M. (1985). The Story of /r/. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (p. 260). Academic Press Inc.

Maddieson, I. (1984). *Patterns of sounds.* Cambridge [Cambridgeshire] ; New York: Cambridge University Press.

Moran, S., & McCloy, D. (Eds.). (2019). *PHOIBLE 2.0.* Jena: Max Planck Institute for the Science of Human History.

Rafat, Y. (2010). A Socio-Phonetic Investigation of Rhotics in Persian. *Iranian Studies*, *43*(5), 667–682.

Sebregts, K. D. C. J. (2014). *The Sociophonetics and Phonology of Dutch r.* Unpublished doctoral dissertation, LOT, Netherlands Graduate School of Linguistics, Utrecht.

Simpson, A. P. (1999). Fundamental problems in comparative phonetics and phonology: does upsid help to solve them. In *Proceedings of the 14th international congress of phonetic sciences* (Vol. 1, pp. 349–352). San Francisco.

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, *12*(1), 1035.

# A COGNITIVE BIAS FOR ZIPFIAN DISTRIBUTIONS? UNIFORM DISTRIBUTIONS BECOME MORE SKEWED VIA CULTURAL TRANSMISSION

Shufaniya, Amir[1], Arnon, Inbal 2*[1]

*Corresponding Author: inbal.arnon@mail.huji.ac.il
[1]Psychology Department, Hebrew University, Jerusalem, ISRAEL

Despite the many differences between them, human languages share certain similarities. One of them is the Zipfian distribution of word frequencies (Zipf, 1949). Across languages, word frequency follows a power law distribution with many low frequency words, few highly frequent words and a non-linear decrease in frequency (e.g., Piantadossi, 2014). The presence of such skewed distributions is a recurring feature of natural language, with ongoing debate about their source and whether they reflect foundational properties of human language/cognition/communication (e.g., Ferrer i Cancho & Sole, 2003). Here, we use an iterated learning paradigm – which shows how weak individual biases impact language structure over time (e.g., Kirby et al. 2008) - to ask whether learners have a cognitive bias for Zipfian distributions, reflected in a tendency to shift uniform distributions into skewed ones during cultural transmission.

Recent work suggests the recurrence of Zpifian distributions in language may be driven (at least in part) by learnability pressures (e.g., Bentz et al., 2017; Lavi-Rotbain & Arnon, 2020, 2022). In particular, the lower unigram entropy of such distributions (compared to uniform and less skewed ones) may benefit learning (Ferrer i Cancho, 2018; Lavi-Rotbain & Arnon, 2022). Indeed, growing evidence suggests Zipfian distributions facilitate word segmentation and learning (Kurumada, Meylan & Frank, 2013; Hendrickson & Perfors, 2019), and that they do so because of their lower unigram entropy (Lavi-Rotbain & Arnon, 2022).

Importantly, existing findings do not tell us whether there is a cognitive preference for skewed distributions: the facilitation may reflect learners' prior experience with such distributions rather than a cognitive preference for them. Here, we explore the existence of such a preference by asking if learners will change their input to make it more skewed. In the first study, we show that learners are biased to produce skewed word distributions in telling a novel story: participants were given a short prompt including six nonce words introduced as names, nouns or verbs (Noun-prompt: "*John has a store with six objects in it: a plizet, a nilbo, a skiger, a vamey, a chila, and a fengle. Write a short story about John's day*"). Aggregated over participants, the nonce words followed a Zipfian distribution for all three parts of speech (high fit of summed frequency by rank to a power law distribution for nouns ($R^2$=.98), verbs ($R^2$=.98), and names ($R^2$=.97).

In the second study, we ask if this bias leads to a shift from uniform distributions towards more skewed ones using an iterated learning paradigm. We exposed the first learner to a story where the six nonce words appeared equally often, and asked them to re-tell it. Their output served as input for the next learner, and so on for a chain of ten learners (or "generations"). We looked at two conditions: (1) where the number of unique words could decrease (five chains) and (2) where it couldn't (participant had to use all words, five chains). Over time, word distributions became more skewed (as measured by lower levels of entropy), even when the number of word types remained constant (significant in model comparisons of mixed-effect regression models, also when excluding differently shaped chain 4). These findings suggest speakers have a cognitive bias for skewed distributions that gets amplified over time, lending support to the idea that their recurrence in language may be driven in part by learnability pressures.

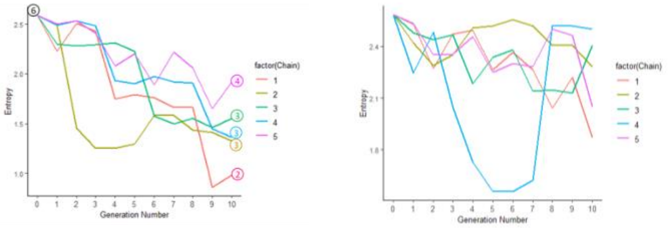


**Figure 3.** The decrease in entropy levels over generations in each of the five diffusion chains in condition 1 (left) and 2 (right)

## References

Ferrer i Cancho, R., & Sole, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences, 100*(3), 788–791.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681-10686.

Kurumada, C., Meylan, S. C., Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition*. *127*, 439–453.

Lavi-Rotbain, O. & Arnon, I. (2022). The learnability consequences of Zipfian distributions, *Cognition, 223, xx-x*

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, *21*(5), 1112–1130. Retrieved from

Zipf, G. K. (1949). Human behavior and the principle of least effort. In *Human behavior and the principle of least effort.* Oxford, England

# NOISE RESISTANCE AT MULTIPLE LEVELS OF LINGUISTIC STRUCTURE

Rachael Bailes*[1], Joel C. Wallenberg[2], and Christine Cuskley[1]

*Corresponding Author: rachael.bailes@ncl.ac.uk
[1]School of English Literature, Language, and Linguistics, Newcastle University, UK
[2]Department of Language and Linguistic Science, University of York, UK

Design for noise resistance is a fundamental requirement of successful communication systems both in technology and the natural world (Shannon, 1948; Brumm & Slabbekoorn, 2005). Redundancy, one way of resisting noise, has been observed at multiple levels of linguistic structure (Fenk-Oczlon, 2001; Aylett & Turk, 2004; Pluymaekers, Ernestus, & Baayen, 2005; Jaeger & Levy, 2007). The ultimate function of noise resistance features is to minimise the loss of *information*, which is quantified in bits for each signal unit in terms of its probability ($log_2 \frac{1}{p_i}$), from Shannon (1948)'s well-known formula for the average information of a set of signals.

We demonstrate that the *ordering* of high and low information units can yield more or less uniform distributions of information, and that more uniform distributions are functional for noise resistance. We also show that speakers are biased toward producing utterances with more uniform distributions (Cuskley, Bailes, & Wallenberg, 2021). Sentences of 10 words were extracted from the *The Penn-York Computer-annotated Corpus of a Large amount of English* (PY-CCLE; 2015), from which we derived four versions of each sentence: Optimised (where the information distribution was made maximally uniform by reordering units with a dedicated algorithm); Asymmetric (where the units were ordered by information content values such that the distribution was entirely asymmetric); Original (the original order of units in the natural sentence), and; Random (where the ordering of units was randomly shuffled). Each set of versions was subjected to two noise conditions: Single Unit Noise (where 3 single units were deleted from the 10 unit string in each version) and Clustered Noise (where 3 sequential units were deleted). The results show that (i) natural sentences were more uniform than random and asymmetric orders, (ii) more uniform distributions lost less information overall, and (iii) uniform distributions of information are 100% successful at preventing *catastrophic information loss* (defined as the loss of $\geq$50% of a distribution's information content). These findings indicate that linguistic units are dynamically ordered across the sentence so as to optimise resistance to noise.

We expand on the theoretical ramifications of an information-loss threshold and the effect of clustered noise as they apply to linguistic communication. Noise (anything that prevents the receiver from observing a complete signal, and only the intended signal) is particularly problematic for language when it comes in a large chunk, due to language's famous "duality of patterning" (Hockett, 1958). Linguistic structure is multi-layered, and the structure at each layer must be inferred from the information a previous layer provides.

An information theoretic approach suggests that each layer of structure serves as a *signal* for another layer. Each layer of structure is a message for which another layer is the signal, and so on. While noise might take a form that affects particular units in isolation (e.g., omitting or altering one particular word in a sentence; (Ryskin, Futrell, Kiran, & Gibson, 2018)), chunks of noise that have the potential to obscure multiple linguistic units are a larger danger from the perspective of catastrophic signalling failures. Even a small amount of noise that spans multiple units - e.g. obscuring the last few phonological segments of one word and the first few of the following word - may well result in a failure to reconstruct *both* words. Because linguistic structure at levels other than the strictly phonetic consists of categorical units, analogue types of noise have the potential to obscure multiple categorical linguistic symbols at once, and so result in knock-on effects disrupting multiple units at higher levels. This may also predict different distributions of information at different levels. Phonemic categories are the first to be signalled, and terminal fading of phonetic material presents a particular noise constraint on this level of signalling. Since information loss compounds, conservation at this first level is crucial. The risk of compounded information loss combined with phonetic processing constraints therefore predict the *frontloading* of phonemic information within a word (King & Wedel, 2020). That phonetic articulation is particularly amenable to redundancy through lengthening (Aylett & Turk, 2004) may serve to mitigate the asymmetry this introduces.

We thus explicate a model of how different levels of linguistic structure signal one another in information theoretic terms. This information-theoretic paradigm for linguistic processing specifies that (i) linguistic communication (i.e. 'message' reconstruction) is a multilayered process of sub-setting the space of possible outcomes based on signal information, and (ii) signals at lower linguistic levels subset the space of possibilities at higher linguistic levels in a serial 'daisy-chain' of inter-level communication, such that (iii) information loss at lower levels of linguistic structure **compounds** information loss at higher, propositional levels.

From this perspective, it follows that linguistic communication is constrained not only by the absolute amount of noise, but by the potential for analogue noise to cause catastrophic signalling failure between linguistic levels. Linguistic planning is biased toward more uniform distributions of information, suggesting adaptation for the prevention of such catastrophic communication failure.

## References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*(1), 31–56.

Brumm, H., & Slabbekoorn, H. (2005). Acoustic communication in noise. *Advances in the Study of Behavior*, *35*, 151–209.

Cuskley, C., Bailes, R., & Wallenberg, J. (2021). Noise resistance in communication: Quantifying uniformity and optimality. *Cognition*, *214*, 104754.

Ecay, A. (2015). *The Penn-York Computer-annotated Corpus of a Large amount of English based on the TCP (PYCCLE-TCP).* Public release 1. https://github.com/uoy-linguistics/pyccle.

Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. *Typological Studies in Language*, *45*, 431–448.

Hockett, C. F. (1958). A course in modern linguistics.

Jaeger, T. F., & Levy, R. P. (2007). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems* (pp. 849–856).

King, A., & Wedel, A. (2020). Greater early disambiguating information for less-probable words: The lexicon is shaped by incremental processing. *Open Mind*, *4*, 1–12.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561–2569.

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, *181*, 141–150.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, *27*(3), 379–423.

# EXPLORING THE CONCEPT OF PHENOTYPIC ACCOMMODATION: THE CASE OF THE LOSS OF AIR SACS

LLUÍS BARCELÓ-COBLIJN[*1]

[*]Corresponding Author: lluis.barcelo@uib.cat
[1]Laboratori d'Investigació en Complexitat i de Lingüística Experimental (LICLE),
Universitat de les Illes Balears, Palma, Spain

During human evolution our ancestors developed a new phenotype that excluded an organ present in all extant great apes: laryngeal air sacs. This change has been acknowledged as an important step towards modern human phonetics. However, to date there is little reflection about how to conceive such a drastic evolutionary change within the subtribe *Hominina.* Here we propose the theoretical integration of air sacs loss as a case of *phenotypic accommodation*, a notion that helps to understand how our ancestors managed to survive, accommodating and consolidating the phenotype without air sacs, paving the way for a new phonetic system.

## 1. Laryngeal air sacs and *Hominoidea*

Laryngeal air sacs is a primitive trait present in many mammals: cetaceans like dolphins and whales, some ungulates (e.g. takins, from the Hymalayas) and also in many primates. Hewitt et al. (2002) showed that, within 128 primate species, up to 72 still have this anatomical feature. Schön Ybarra (1995) attested up to four different kinds of laryngeal air sacs within primates: (1) lateral ventricular, (2) subhyoid, (3) infraglottal and (4) dorsal. The first one is present in all great apes with the exception of *H. sapiens*. Air sacs are present in juvenile apes, even in fetuses and neonates (Stark & Schneider 1960). Steele et al. (2013) used 3-D reconstructions and observed that air sacs in chimpanzees are lateral ventricular, extending in a position behind the clavicle bone.

### *1.2. Possible functions of air sacs*

The relation of the vocal tract and air sacs and their potential functions is controversial, since their potential functions could be several and not all species use air sacs for the same functions: for saving exhaled air (Negus, 1949); for the reduction of hyper-ventilation (Hewitt et al., 2002); for generating a new sound source and stronger and longer lasting calls (Fitch & Hauser, 2003). Lieberman (2011) notes that air sacs are probably related to vocalization since inspiring carbon dioxide-rich air is not as useful as controlling speech. Falk (1975) pointed out that the vertical movement of the hyoid bone also compresses the orifice of the laryngeal air sacs and hence, participating in the mechanics when air sacs are emptied out and filled up. De Boer (2008) and Riede et al. (2009) created several models showing the influence of the sound produced by air sacs on the sound produced by vocal folds, while experimental research shows that, when both kind of sounds become superimposed, modern humans have difficulties in order to distinguish vowels properly (de Boer, 2012). Although Harrison (1995) dismissed the idea that air sacs are not necessary for vocalizations, several scholars do not agree. Lieberman (2010: 333) recalls as well that another function of air sacs in gorillas is acting "like a resonating drum" when gorillas thump their chest. More recently, Perlman & Salmi (2017) analyzed the vocalizations of gorillas and suggest that this species could use them for male display.

### *1.3. Abrupt change and the view of complexity*

During the evolution of the subtribe *Hominina*, laryngeal air sacs disappeared at some point. The key feature to discern the presence or absence of air sacs in the fossil record seems to be the morphology of hyoid bone. The australopithecine (*Au. afarensis*) hyoid bone found, by Alemseged et al. (2006) in Dikika (Ethiopia), shows a primitive morphology that would suggest the connection of air sacs to the respiratory system and, hence, to the whole apparatus in charge of speech. The shape of the Dikika hyoid shows a deep bulla, while the shape of the *H. sapiens* hyoid describes an arch, without that deep cavity on the body section. The deep bulla would be a result of the development of the individual, from the direct contact of the tissues of air sacs to the hyoid bone.

The next fossilized hyoid bone from the fossil record seemed to belong to a *H. erectus* (Capasso et al., 2008). That particular hyoid bone showed a modern morphology, i.e. there is no cavity or bulla shape to which air sacs could be in contact with. However, a second examination by Capasso et al. (2016) confirmed that that bone was "too thick and short to be the body of human hyoid". Finally, the neanderthal hyoid bone is quite similar to the *H. sapiens'* hyoid bone (Arensburg et al., 1989; Bar-Yosef et al., 1992).

Thus, at some temporary point of the early stages of genus *Homo* – somewhere between the *Australopithecus* and *H. erectus* clades–, an early hominin started making steps into modern speech. A plausible, falsifiable hypothesis is that *H. habilis* could be that hominin. Suggestions about *H. habilis* capability for speech are not new at all (Tobias, 1987), but information about this hominin is still too scarce and contentious (Wood & Collard, 1999) and, hence, this hypothesis is, for the moment, still more grounded upon plausible ideas than upon solid data from the fossil record.

### *1.4. Modularity and complexity*

The enigma of air sacs loss in our lineage needs an explanation integrating the fact that the tissue of an air sac is always in physical contact with other tissues: hard tissues (the hyoid bone) and soft tissues (muscles and ligaments). This fact establishes a direct relationship between them, since all tissues interact with each other, not only due to contact, but also through the movements of the body (e.g. movements of the laryngeal muscles when producing sounds or during deglutition; or while turning the head). In fact, this can be considered as a collection of interacting elements or, in other words, a complex system. In such systems, the deletion of some well-connected elements leads to a reorganization of the system.

The structure of organisms seem to present a modular building schema (Wagner, 1996; Schlosser & Wagner, 2004; Callebaut & Rasskin-Gutman, 2005). Heads have been analyzed as a complex system which integrates several modules and submodules, which include different sets of bones (more than 20), teeth (up to 32), a brain, sensory organs, muscles, ligaments, cartilages, veins, nerves, etc., showing altogether an evident modular organization (Lieberman, 2011: 8-12). Heads include the neck and all its hard and soft tissues. Inside the human neck two passages can be detected, one for the air and another for the food. Anatomical network analyses have revealed musculoskeletal modularity in primates (Esteve-Altava et al., 2015a; Esteve-Altava et al., 2015b; Powell et al., 2018).

In a recent network analysis of musculoskeletal organization, Powell et al. (2018) show that, in spite of having more musculoskeletal elements (up to 157), humans do not seem to be more complex than other great apes. These authors argue that “even major changes in function […] can occur without profound changes to the network organization and modularity of the whole system”. Although Powell and collaborators include only direct musculoskeletal connections and not organs like air sacs, this reflection is still valid for the discussion on the evolution of air sacs: air sacs can disappear, as it has occurred in many taxa, without compromising the network organization, but affecting the function of the module. Following Powell et al. (2018), developmental plasticity can facilitate the accommodation of functional and anatomical modifications without conditioning severely the network configuration.

Being it as it may, when an individual develops following a divergent and atypical developmental path, for example lacking (part of) an organ, the unexpected phenotype can provoke that other tissues of the system develop abnormally as well, due to the absence of physical opposition, and due to the "departure" –metaphorically speaking– from what was "planned" by the original, species-specific, body schema. Hence, soft tissues interact with hard tissues through physical contact. Lieberman recalls a good example of it in humans: an infant born lacking eyes will probably develop "tiny orbits with abnormally small upper faces because the eyeball normally pushes out the bones that form the orbital cavity walls" (2011: 53). Similarly, the air sacs loss had to have consequences for the surrounding tissues.

Other forces, like persistent muscular movements, if lasted through generations, could also exert an influence on the phenotypes, which could have been finally selected. Far from simply resorting to Lamarck's (1809) *Zoological Philosophy*, we advocate a hypothesis in which phenotypes, not mutations, are the selected ones (West-Eberhard, 2003). For example, recent analyses of dry skulls of both humans and chimpanzees suggest a link between the large number of mandibular movements of muscles involved in speech (much larger than those used in chewing) and the morphology of the mandibular space in relation to the tongue (Bermejo-Fenoll et al. 2019). In spite of the difficulty for showing a cause-effect relationship, Bermejo-Fenoll and colleagues could show that *H. sapiens*' ancestors followed an evolutionary trend which, progressively and increasingly, included more and more movements that today are related to modern speech, like for example lip smacking and oscillatory movements of the jaw (Ghanzafar et al., 2013; Ghanzafar et al., 2012; MacNeilage & Davis, 2005). Anyway, it seems out of question that, at some point between australopithecines and modern humans and –we suggest– after the disappearance of air sacs, the physical context of speech and the laryngeal structure changed, paving the way to new motor routines which could involve more (and perhaps new) movements that today are related to modern speech.

## 2. Air sacs loss as a case of phenotypic accommodation

West-Eberhard (2005a: 6547) recalls that "frequency of expression does not depend on the frequency of the inducer (mutation or environmental factor) alone". Thus, according to West-Eberhard's (2003, 2005b) proposal, evolutionary selection combines both natural and sexual or social selection, and hence, genes would be followers, not leaders in adaptive evolution. We think this is a good theoretical framework for a change like the loss of air sacs. This is something that has not been yet approached and, thus, there are some possibilities open to inquiry. One possibility is that the first individuals lacking air sacs represented atypical phenotypes within their group of conspecifics. Hence, it is expectable that the disappearance of air sacs from the body schema altered, even if slightly, some developmental patterns during ontogeny, and thus yielding new, different phenotypes in adult individuals. Some available evidence from comparative studies do support such a view: works on the ontogeny of the descent of the hyolaryngeal complex and the root of the tongue takes place at about 3 months of age (Lieberman et al., 2001). By contrast, at 4 months of age the initial pouch that will become the air sac is already visible in chimps (Nishimura et al., 2007).

The observable fact is that this new phenotype somehow reached stabilization, and one of the reasons could be that it was not incompatible with survival. This can be interpreted as an indication (though still not confirmed evidence) of a fitness effect. According to West-Eberhard (2005a), if the new phenotype has a fitness effect, then selection occurs. Were this the case of air sacs loss, it could be conceived as a phenotype that has been consolidated, becoming the typical phenotype of current *H. sapiens*. Observed through this lens, thus, the air sacs loss could be classified a case of *phenotypic accommodation* (West-Eberhard, 2003; 2005b). Phenotypic accommodation takes place when an individual develops plastically, adapting the organism to new environmental inputs (here "environmental" refers to all levels in biology, from cells to ecological niches). If the new inputs persist and other conspecifics can develop the same adaptive phenotype as well, this new phenotype could become stabilized by, for example, new mutations promoting and reinforcing it.

Arguments supporting this hypothesis are, for example, that air sac loss is a consolidated phenotype in many primate species, some of them phylogenetically distant from one another (cf. Hewitt et al., 2002). Thus, it seems to be a recurrent homoplasy within the *Primates* order. Recurrent phenotypes with discontinuous phylogenetic distributions have been detected in many taxa (cf. West-Eberhard, 2003, for a highly detailed revision). Moreover, the existence of both phenotypes within *Primates* suggests ancestral developmental plasticity for producing both forms (West-Eberhard, 2005b: 6546). Were the air sacs loss phenotype dependent of a mutation (or methylation) only, it would be a mutation/methylation that has appeared many times and hence, less plausible. Thus, phenotypic plasticity seems to be a good candidate for the origins of air sac loss, given the fact that it can account for the repetition of homoplasies in different primates and in different stages of phylogeny. Moreover, phenotypic accommodation could give time to this developmental variant for becoming stabilized within populations and species. The absence of air sacs would have led the possibility for some individuals to develop towards new phenotypes that were compatible with life (here we follow Pere Alberch's (1989) "logic of monsters" and assume the idea of a phenotypic space of biological possibilities). Let us remark that this hypothesis is conceived as a phenotypic possibility for some ancestors of *H. sapiens*, and not for all species that have evolved towards phenotypes lacking air sacs. The reason is that, usually, there is not a unique factor affecting the several environmental levels by which an organism is affected, and each species is affected by a different set of environmental factors (admittedly, partially coincidental with other species). The first individuals who developed like that, were at first unusual and atypical. However, some of their descendants managed to survive, consolidating (accommodating) a phenotype without air sacs, and thus paving the way for a new phonetic system.

## Conclusion

Laryngeal air sacs is an ancestral trait in *Primates*, while its absence is a derived characteristic. Both possibilities exist in monkeys, lesser apes (*Hylobatidae*) and great apes (*Hominidae*). It is well-known that the descent of larynx does not guarantee vocalization (Fitch 2009). Similarly, the lack of air sacs does not guarantee neither the descent of larynx nor modern speech. But it was a necessary step for evolving towards what is known today as modern human speech. Nonetheless, both phenotypes are recurrent and it is puzzling how to integrate this into evolutionary theory. We propose that this could be a case of phenotypic plasticity and, more concretely, a case of phenotypic accommodation. This concept satisfies the recurrence of both phenotypes and the differences in ontogeny in juvenile chimps and humans. Finally, the concept of phenotypic accommodation solves the difficulty of linking this evolutionary change to mutation only, and gives the time a species needs until the new variant becomes stabilized.

## Acknowledgements

This research was partially supported by grant number PID2021-128404NA-I00 from the Ministerio de Ciencia e Innovación (Spain).

## References

Alberch, P. (1989). The logic of monsters: Evidence for internal constraint in development and evolution. *Geobios*, 22(2): 21-57.

Alemseged, Z., F. Spoor, W. H. Kimbel et al. (2006). A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature,* 443:296–301.

Arensburg, B., Tillier, A-M., Vandermeersch, B., Duday, H., Schepartz, L.A. & R. Yoel. (1989). A Middle Palaeolithic human hyoidbone. *Nature* 338, 758–760.

Bar-Yosef, O., Vandermeersch, B., Arensburg, A., Belfer-Cohen, P., et al. (1992). The excavations in Kebara Cave, Mt. Carmel. *Current Anthropology*, 33: 497–550.

Bermejo-Fenoll, A., Panchón-Ruiz, A., Sánchez del Campo, F. (2019). *Homo sapiens*, chimpanzees and the enigma of language. *Frontiers in Neuroscience*, 13(558): 1-10.

de Boer B. (2012). Loss of air sacs improved hominin speech abilities. *Journal of Human Evolution*, 62(1):1-6.

de Boer, B. (2008). The Joy of Sacs. In A. Smith, K. Smith, & R. Ferrer i Cancho, (Eds.), *The Evolution of Language: Proceedings o the 7th international Conference (EVOLANG7)*, 415–416. Barcelona.

Callebaut, W., & Rasskin-Gutman, D. (2005). *Modularity. Understanding the development andEvolution of Natural Complex Systems*. Cambridge, MA: MIT Press.

Capasso, L., Michetti, E. & D'Anastasio, R. (2008). A *Homo erectus* hyoid bone: possible implications for the origin of the human capability for speech. *Collegium Antropologicum,* 32 (4): 1007-1011.

Capasso, L., D'Anastasio, R., Mancini, L., Tuniz, C. & Frayer, D. (2016). New evaluation of the Castel di Guido 'hyoid'. *Journal of Anthropological Science*, 20;94:231-5.

Esteve-Altava, B., Diogo, R., Smith, C., Boughner, J. C., & Rasskin-Gutman, D. (2015). Anatomical networks reveal the musculoskeletal modularity of the human head. *Scientific Reports*, *5*(1): 1-6.

Esteve-Altava, B., Boughner, J. C., Diogo, R., Villmoare, B. A., & Rasskin-Gutman, D. (2015). Anatomical network analysis shows decoupling of modular lability and complexity in the evolution of the primate skull. *PLoS One*, 10(5): e0127653.

Falk, D. (1975). Comparative anatomy of the larynx in man and the chimpanzee: Implications for language in Neanderthal. *American Journal of Physical Anthropology*. 43:123–132.

Fitch W.T. and Hauser M.D. (2003) Unpacking "Honesty": Vertebrate Vocal Production and the Evolution of Acoustic Signals. In: *Acoustic Communication*. Springer Handbook of Auditory Research, vol 16. New York: Springer.

Fitch, W.T. (2009). Fossil cues to the evolution of speech. In: Botha, R., Knight, C. (Eds.), *The Cradle of Language*. Oxford University Press, Oxford, pp. 112–134.

Ghazanfar, A.A., Takahashi, D.Y., Mathur, N., Fitch, W.T., (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology*, 22 (13), 1176–1182.

Ghazanfar, A.A., Morrill, R.J., Kayser, C., (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Science,* 110 (5), 1959–1963.

Harrison, D. F. N. (1995) Detailed morphology. In *The anatomy and physiology of the mammalian larynx*. 48-184. Cambridge: Cambridge University Press.

Hewitt, G., MacLarnon, A. & K. E. Jones. (2002). The functions of laryngeal air sac in primates: Anew hypothesis. *Folia Primatologica*. 73:70–94.

Lamarck, J. B. (1809; 1914). *Zoological Philosophy*. London.

Lieberman, D.E. (2011). *The Evolution of the Human Head*. Cambridge, MA: Harvard University Press.

Lieberman, D.E., McCarthy, R.C., Hiiemae, K.M. & Palmer, J.B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*, 46: 117-128.

MacNeilage, P.F. & B.L., Davis. (2005). The frame/content theory of evolution of speech: a comparison with a gestural-origins alternative. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 6 (2), 73–99.

Negus, V. E. (1949). *The comparative anatomy and physiology of the larynx*. 1st ed. W. Heinemann Medical Books.

Nishimura, T., A. Mikami, J. Suzuki et al. (2007). Development of the laryngeal air sac in chimpanzees. *International Journal of Primatology*. 28:483–492.

Perlman, M., & R. Salmi. (2017). Gorillas may use their laryngeal air sacs for whinny-type vocalizations and male display. *Journal of Language Evolution*, 2(2): 126–140,

Powell, V., Esteve-Altava, B., Molnar, J. et al. (2018) Primate modularity and evolution: first anatomical network analysis of primate head and neck musculoskeletal system. *Scientific Reports,* 8: 2341.

Riede T., Tokuda I.T., Munger J.B., Thomson S.L. (2008). Mammalian laryngseal air sacs add variability to the vocal tract impedance: physical and computational modeling. *Journal of the Acoustical Society of America*,124(1):634-47.

Schön Ybarra, M. (1995). A comparative approach to the nonhuman primate vocal tract: Implicationsfor sound production. In *Current Topics in Primate Vocal Communication*, E. Zimmerman and J. D. Newman, eds. New York: Plenum Press, 185–198.

Starck, D. & Schneider, R. (1960). Respirationsorgane. In H. Hofer, A. H. Schultz & D. Starck (Eds.), *Primatologia* (Vol. III (2), pp. 423–523). Basel: S. Karger.

Steele J, Clegg M, Martelli S. (2013). Comparative morphology of the hominin and African ape hyoid bone, a possible marker of the evolution of speech. *Human Biology*, 85(5):639-72.

Tobias, P.V. (1987). The brain of Homo habilis: A new level of organization in cerebral evolution. *Journal of Human Evolution*, 16(7–8): 741-761.

Schlosser, G., & Wagner, G. P. (Eds.). (2004). *Modularity in development and evolution*. University of Chicago Press.

Wagner, G. P. (1996). Homologues, Natural Kinds and the Evolution of Modularity. *American Zoologist*, 36:36-43.

West-Eberhard, M.J. (2005a). Developmental plasticity and the origin of species origin. *Proceedings of the National Academy of Science*, 102(1): 6543-6549.

West-Eberhard, M.J. (2005b). Phenotypic accommodation: adaptive innovation due to developmental plasticity. *Journal of experimental zoology. part B, (Mol Dev Evol)*, 15;304(6):610-8.

West-Eberhard, M.J. (2003). *Developmental Plasticity and Evolution*. New York: Oxford Univ. Press.

Wood, B. & M. Collard. (1999). The human genus. *Science*, 284, 65–71.

# PREVALENCE BASED QUANTITATIVE ANALYSIS OF INNOVATIONS IN LEXICAL EVOLUTION

ANDREAS BAUMANN*[1] and KLAUS HOFMANN[1]

*Corresponding Author: andreas.baumann@univie.ac.at
[1]Faculty of Philological and Cultural Studies, University of Vienna, Vienna, Austria

The spread of linguistic innovations has been modeled by means of logistic growth (Altmann, 1985; Labov, 1994; Ghanbarnejad et al., 2014). The epidemiological argument underlying this type of growth is that (a) innovations are transmitted from one individual to the next and that (b) growth rate depends on both the number of individuals that already know the innovation and the number of individuals that do not yet know it (Figure 1b, left; Cavalli-Sforza & Feldman, 1981). The resulting dynamics are S-shaped (as in Figure 1c; cf. Blythe & Croft, 2012). Interestingly, although this argument is inherently population dynamic and involves interacting individuals, empirical accounts typically assess S-shaped linguistic spread by means of token frequency (Denison, 2003).

In this contribution, we promote the use of *prevalence* for studying lexical change, i.e., the fraction of individuals that know and use a word. Lexical prevalence was employed in mathematical accounts of lexical spread (Nowak, 2000) and, more recently, considered in psycholinguistic research where prevalence was argued to function as relevant (control-)variable in experimental setups (Brysbaert et al., 2019). Here, we show (a) that lexical prevalence is relevant for studying the spread of lexical innovations in that prevalence trajectories typically precede frequency trajectories, and we propose (b) the use of coupled dynamics of prevalence and frequency. Our study unfolds in two steps.

In the first part, we use a diachronic corpus of German newspapers spanning a period of two decades (Ransmayr et al., 2017) and derive trajectories of token frequency and prevalence (by using the fraction of authors using a word as a proxy for prevalence; cf. Johns et al., 2020) for a set of about 700 words that have been strongly increasing in this period (e.g., *googeln*, ‘to google’). By fitting logistic models to each trajectory, we determine (i) the intrinsic growth rate, (ii) the inflection point, and (iii) goodness-of-fit for each word. We show that prevalence curves have (i) a higher intrinsic growth rate (cf. Figure 1a), (ii) earlier inflection points, and (iii) a better fit that frequency curves. We take this to show that logistic prevalence curves detect lexical innovations earlier and more reliably than this is the case for logistic frequency curves.

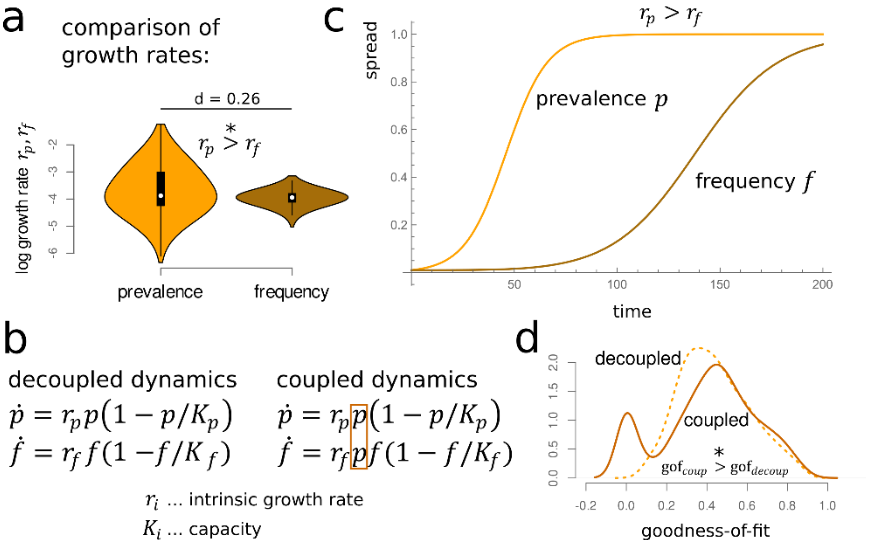


Figure 1. (a) Empirical comparison of prevalence and frequency trajectories; (b) studied dynamical systems; (c) solution for coupled dynamics; comparison of goodness-of-fit of the systems in (b).

In the second part of our study, we test if goodness-of-fit improves if we couple the dynamics of prevalence and frequency rather than keeping them separate as before. This is done by letting the growth of frequency at a certain time not only depend on token frequency but also on prevalence (Figure 1b, left). This is motivated by the assumption that the production of words depends on whether an individual knows a word and on how entrenched it is (Bybee, 2006). Given that intrinsic growth is higher for prevalence than for frequency, the frequency trajectory is shown to be preceded by the prevalence trajectory in this model (Figure 1c). Subsequently, we use the empirical trajectories analyzed before and compute goodness-of-fit for the respective solutions of both systems, coupled and decoupled. We show that although coupled dynamics are in some cases uninformative (goodness-of-fit close to zero), they generally show a slightly better fit than decoupled systems (Figure 1d; our study is complemented with a sensitivity analysis of goodness-of-fit differences). We conclude that coupled dynamics of prevalence and frequency represent a reasonable model for studying lexical evolution that, in addition, help to reconstruct diachronic prevalence trajectories from frequency data.

## References

Altmann, G. (1985). On the dynamic approach to language. In *Linguistic Dynamics* (pp. 181-189). de Gruyter.

Blythe, R. A., & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 269-304.

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior research methods*, *51*(2), 467-479.

Bybee, J. (2006). *Frequency of use and the organization of language*. Oxford University Press.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton University Press.

Denison, D. (2003). Log (ist) ic and simplistic S-curves. *Motives for language change*, *54*, 70.

Ghanbarnejad, F., Gerlach, M., Miotto, J. M., & Altmann, E. G. (2014). Extracting information from S-curves of language change. *Journal of The Royal Society Interface*, *11*(101), 20141044.

Johns, B. T., Dye, M., & Jones, M. N. (2020). Estimating the prevalence and diversity of words in written language. *Quarterly Journal of Experimental Psychology*, *73*(6), 841-855.

Nowak, M. A. (2000). The basic reproductive ratio of a word, the maximum size of a lexicon. *Journal of theoretical biology*, *204*(2), 179-189.

Ransmayr, J., Mörth, K., & Ďurčo, M. (2017). AMC (Austrian Media Corpus) – Korpusbasierte Forschungen zum österreichischen Deutsch. In Claudia Resch and Wolfgang U. Dressler (eds.), *Digitale Methoden der Korpusforschung in Österreich*. 27-38. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

# LESS OF THE SAME: MODELING *HORROR AEQUI* AND EXTRAVAGANCE AS MECHANISMS OF NEGATIVE FREQUENCY DEPENDENCE IN LINGUISTIC DIVERSIFICATION

ANDREAS BAUMANN[*1] and ROLAND MÜHLENBERND[2]

[*]Corresponding Author: andreas.baumann@univie.ac.at
[1]Faculty of Philological and Cultural Studies, University of Vienna, Vienna, Austria
[2]Leibniz-Centre General Linguistics, Berlin, Germany

Two mechanisms involving negative frequency dependence, i.e., pressures for less of the same, have been suggested to affect linguistic evolution: (i) *horror aequi*, a cognitive-psychophysiological effect that leads to avoidance of repetitions, and (ii) extravagance, which is a pragmatic strategy to signal non-conformity and emphasis by avoiding predominant constituents in favor of innovative ones. In this paper, we explore how avoidance biases like this influence the long-term evolution of linguistic constituents. We do so by means of evolutionary analysis of population dynamic models of linguistic diffusion. We show that if individuals are sufficiently sensitive with respect to variation then avoidance biases can yield linguistic diversification, i.e., stable coexistence of constituent variants.

## 1. Introduction

Linguistic systems exhibit diversity on various levels. Phonological systems consist of diverse sounds, some of which are closely related to each other (e.g., more or less leniated variants of a consonant like /b/). Vocabularies usually consist of thousands of words, some of which may be formally similar but semantically different (e.g., *freak* denoting an abnormal individual or, in fact, an expert or aficionado). Other words are formally different but denote the same concept (e.g., *session* and its slang variant *sesh*). And on the syntactic level, languages feature a multitude of constructions, although some of them are formally slightly different but might fulfill the same function (e.g., [*going to* + V] and [*gonna* + V] for expressing future events), or indeed formally similar but semantically different (e.g., [*going to* + V] and [*going to* + N]). Where does this diversity come from?

One of the potential causes of the successful implementation of variants like the ones above is the tendency to avoid more of the same. We will refer to this as *avoidance bias*. Two mechanisms have been discussed in this matter. On the cognitive-psychophysiological level, the *horror aequi* effect was suggested to promote variation in that it provides a bias against repetitions. On the pragmatic level, *extravagance* was discussed as a mechanism that enhances innovations to make their users stand out of the crowd (Haspelmath, 2000; Walter, 2007; Petré, 2017; Ungerer & Hartmann, 2020). In this paper, we study the effect of such biases on linguistic diversity by means of population-dynamic models. For this, we implement frequency dependent avoidance biases into a dynamical system that characterizes the spread of a linguistic constituent (phoneme, word, construction) through a population of speakers and analyze the long-term evolution of this constituent (Doebeli & Ispolatov, 2010; Dercole & Rinaldi, 2008). We show that diversification of the constituent, i.e., the stable establishment of two variants of itself, depends on the scope of the mechanism and how easy it is for speakers to differentiate between variants.[1]

The subsequent section provides a brief review of the two mechanisms, *horror aequi* and extravagance. After that, we outline our model together with its evolutionary analysis and finally discuss its implications and limitations.

## 2. *Horror aequi* and extravagance in linguistic evolution

Avoidance biases can operate on different levels. The *horror aequi* effect ('fear of the same') was suggested as a mechanism that lets speakers avoid repetitions of similar (or even identical) linguistic structures operating on various levels of linguistic organization. The effect has been suggested to be motivated by psychological and physiological constraints as well as constraints on cognitive planning (Walter, 2007). In the phonological domain, it was proposed as a mechanism for avoiding repetitions by Brugmann (1917) already in the beginning of the 20th century (e.g., German *Zauberin* rather than *Zaubererin*, where *Zauberer* is a 'male magician' and *-in* is the German feminine suffix). In phonetics and phonology, avoidance biases have been studied under the term *obligatory contour principle*. For instance, Walter (2008) has shown that repetition of the consonants /b, d, g/ leads to the production of lenited variants. In the morpho-syntactic domain, the effect was studied by Rohdenburg and others to account for the avoidance of repeating *-ing* forms in English in sequences like *without bothering to tell him* vs. *without bothering telling him* (Szmrecsanyi, 2008; Rohdenburg, 2011). The effect typically applies within utterances but it can also apply across utterance boundaries, as long as they are temporally close. Indeed, *horror aequi* was suggested to be a potential counterpart of (asymmetric) priming (Jäger & Rosenbach, 2008) and attested experimentally (Hilpert & Saavedra, 2018) (although *horror aequi* was argued by Szmrecsanyi (2008) to be weaker than supportive effects).

Extravagance, in contrast, is a pragmatic phenomenon that by definition applies to interactions among individuals. In an excellent recent review, Ungerer and Hartmann (2020) elaborate on the different aspects of extravagance covered

---

[1] In biological evolution, negative frequency-dependent selection is a well-studied mechanism potentially accounting for diversification and stable polymorphism (Brisson, 2018; Chesson, 2000). It is less well studied in cultural evolution; but see, e.g., Doebeli and Ispolatov (2010), who use a model similar to that in this paper to model coexistence of religions.

in the literature. An important aspect of extravagant language seems to be that it lets their users stand out of the crowd. While this may have the reason to signal (out-)group membership (Fajardo, 2019) in a non-conformative fashion (e.g., back-clipping *session* to *sesh* in youth slang), this is not necessarily always the case. Extravagant expressions may be as well used to gain attention or as emphasis device. So, the [*going to* + V] future construction or the progressive construction [*be* V-*ing*] were argued to be motivated by extravagance, as was the lengthened an more explicit expression *by means of* in place of *with* (Haspelmath, 1999, 2000; Vosberg, 2003; Petré, 2017). Length seems to be a particularly relevant property related to extravagance, as seen in phrasal compounds like *make-your-stomach-hurt difficult* (Günther, Kotowski, & Plag, 2020).

What both mechanisms have in common is (i) the tendency to avoid a predominant variant in favor of another variant and (ii) frequency dependence. That is, both *horror aequi* and extravagance only apply if certain structures are used and encountered relatively often. *Horror aequi* requires nearby repetitions of linguistic structures and extravagant expressions require a background of predominantly used forms against which they are perceived as extravagant in the first place. Frequency dependence is negative because avoidance biases impede the usage of the predominant variant. In what follows, we model avoidance biases and their effect on linguistic diffusion through speaker populations.

## 3. Modeling linguistic evolution under avoidance biases

Let us consider a linguistic constituent that is characterized by a certain property $x$. Such a constituent could be, e.g., a phoneme, a phoneme sequence, a word, or a construction, and its property $x$ could be some formal aspect (such as the degree of lenition of the consonant /b/) or some semantic aspect (like the sentiment of the word *freak*, which could be positive, negative, or something in between). We assume that $x$ can be measured on a continuous scale, i.e., $x \in \mathbb{X} \subseteq \mathbb{R}$, where $\mathbb{X}$ is an interval defining the range of possible values. By learning the constituent, it is transmitted from one individual to the next, either horizontally, or vertically in first-language acquisition. Whenever a user of the constituent and a learner who does not yet know it meet, successful learning takes place at an intrinsic rate $\lambda$ depending on $x$. We also assume that learning is optimal for some value $x_0$ (i.e., $f$ is locally concave around $x_0$). When an individual learns the constituent, they switch from the learner to the user class. Let $L$ and $U$ denote the respective number of individuals.

Individuals can also cease to use a constituent so that they switch from the user class back to the learner class. In our model, switching back to the learner class is motivated by avoidance biases outlined before. Thus, users cease to use a constituent whenever they interact with another user of that constituent at an avoidance rate $\alpha$. That is, growth is subject to negative frequency dependence (Brisson, 2018). Oftentimes, individuals using slightly different variants will interact. Suppose that the values $x$ and $y$ define two different variants of a constituent (two different degrees of lenition; two different sentiment values). We assume that $\alpha$ is a decreasing function of the distance $\Delta = |x - y|$ between both variants obtaining its maximum $\alpha_0$ at $\Delta = 0$. Avoidance rate $\alpha$ is highest if both variants are identical and decreases the more different they are. A function that models such a behavior is

$$\alpha(\Delta) = \alpha_0 \exp(-1/2 \cdot \Delta^2/\sigma^2). \tag{1}$$

This is a bell-shaped curve depending on $\Delta$. The steepness of this curve is determined by the parameter $\sigma$, which stands for the *scope* of the avoidance mechanism. The scope defines the range of values that are affected by that mechanism in a fuzzy manner (Figure 1a, top panel). If $\sigma$ is high (flat curve) then even substantially different variants lead to high mutual avoidance rates. If $\sigma$ is low (steep curve) then only relatively similar constituents will lead to high avoidance rates. The reciprocal $1/\sigma$ can be interpreted as a measure of the *sensitivity* with respect to variation. High sensitivity (low $\sigma$) means that only nearby variants are perceived as identical. Low sensitivity (high $\sigma$) means that most variants are perceived as identical, which promotes avoidance. Note that, for simplicity, our model only captures mechanisms of negative frequency dependence although language clearly shows positive frequency dependence as well; see Doebeli (2011) for a model juxtaposing conformity and non-conformity biases. We will come back to this limitation in the final section.

Given the above considerations, the population dynamics of a constituent specified by $x$, in the absence of other variants, is given by the differential equation

$$\dot{U} = \lambda(x)UL - \alpha_0 UU, \tag{2}$$

where we assume that population size is normalized so that $L + U = 1$. If $\lambda(x) > 0$ (positive learning rate) the population dynamic equilibrium is given by $\hat{U}(x) = \lambda(x)/(\lambda(x) + \alpha_0)$.

What can we say about the long-term evolution of the constituent's property $x$? We can model the evolutionary trajectory of $x$ with the help of the canonical equation of adaptive dynamics (Meszena, Kisdi, Dieckmann, Geritz, & Metz, 2002; Dercole & Rinaldi, 2008), which relates to the Price equation (Page & Nowak, 2002) and defines the rate of change of $x$ as

$$\dot{x} = M \frac{\Sigma^2}{2} \hat{U}(x) \left. \frac{\partial f(x,y)}{\partial y} \right|_{y=x}. \tag{3}$$

Here, the constants $M$ and $\Sigma^2$ define the rate and variance of linguistic innovations, respectively. More importantly, $f(x, y)$ denotes invasion fitness, i.e., the exponential growth rate of a rare variant characterized by $y$ in a population in which $x$ is the predominant variant. That is, for every $x$, $f(x, y)$ defines a fitness landscape that rare variants $y$ have to cope with (Figure 1a, mid panel). Given the population dynamics in (2), invasion fitness can be derived as

$$f(x, y) = \lambda(y)(1 - \hat{U}(x)) - \alpha(y - x)\hat{U}(x). \tag{4}$$

One can show that $x_0$ (where learning is optimal) defines an equilibrium of the evolutionary dynamics (3) (Doebeli & Ispolatov, 2010). Moreover, $x_0$ is an evolutionary attractor, so that values close to $x_0$ are driven towards this equilibrium.

What is more interesting is this: if $\sigma$ is large (low sensitivity with respect to variation) then $x_0$ is a local maximum of the fitness landscape given by $f(x_0, y)$ (Figure 1a, bottom panel). This can be seen by looking at the curvature of $f$ around $x_0$ (because $\alpha''(0)$ is small for large $\sigma$):

$$\left.\frac{\partial^2 f(y, x_0)}{\partial y^2}\right|_{y=x_0} = \lambda''(x_0)(1 - \hat{U}(x_0)) - \alpha''(0)\hat{U}(x_0). \tag{5}$$

This means that $x_0$ cannot be invaded by nearby variants. If, however, $\sigma$ is sufficiently small then $x_0$ is a local minimum of the fitness landscape $f(x_0, y)$, so that $x_0$ can be invaded by nearby variants. Thus, $x_0$ is a *branching point* at which the population is split into two variants that stably coexist (Geritz, Metz, Kisdi, & Meszéna, 1997). This is shown in Figure 1b. High sensitivity with respect to variation combined with avoidance biases leads to diversification of the constituent into two variants (e.g., more/less lenited; positive/negative).

## 4. Discussion and conclusion

We have seen that mechanisms accounting for negative frequency dependence can drive linguistic diversification. Two such mechanisms have been discussed: (i) the *horror aequi* effect operating on the cognitive-psychophysiological level and (ii) extravagance, i.e., the tendency to behave differently in order to stand out as a pragmatic phenomenon. In the literature, both mechanisms have been suggested to promote linguistic innovations and hence linguistic evolution. What we have shown in our contribution, though, is that stable coexistence of two variants (i.e., diversity) that result from one of these mechanisms depends on how sensitive individuals are with respect to variation. If individuals do not differentiate between different variants and put all of them into one basket then neither *horror aequi* nor extravagance will lead to diversification. Both mechanisms require a certain ability to differentiate between variants. Without this ability, linguistic evolution simply optimizes learnability (so that, say, only lenited /b/ and *freak* with a positive connotation would remain).

Why is this plausible? If (i) individuals treat almost all variants as identical then *horror aequi* is very likely to apply. If, however, *horror aequi* only applies to very similar variants, the mechanism will lead to deletion much more rarely, hence enforcing coexistence. Likewise, if (ii) even distant variants are considered the

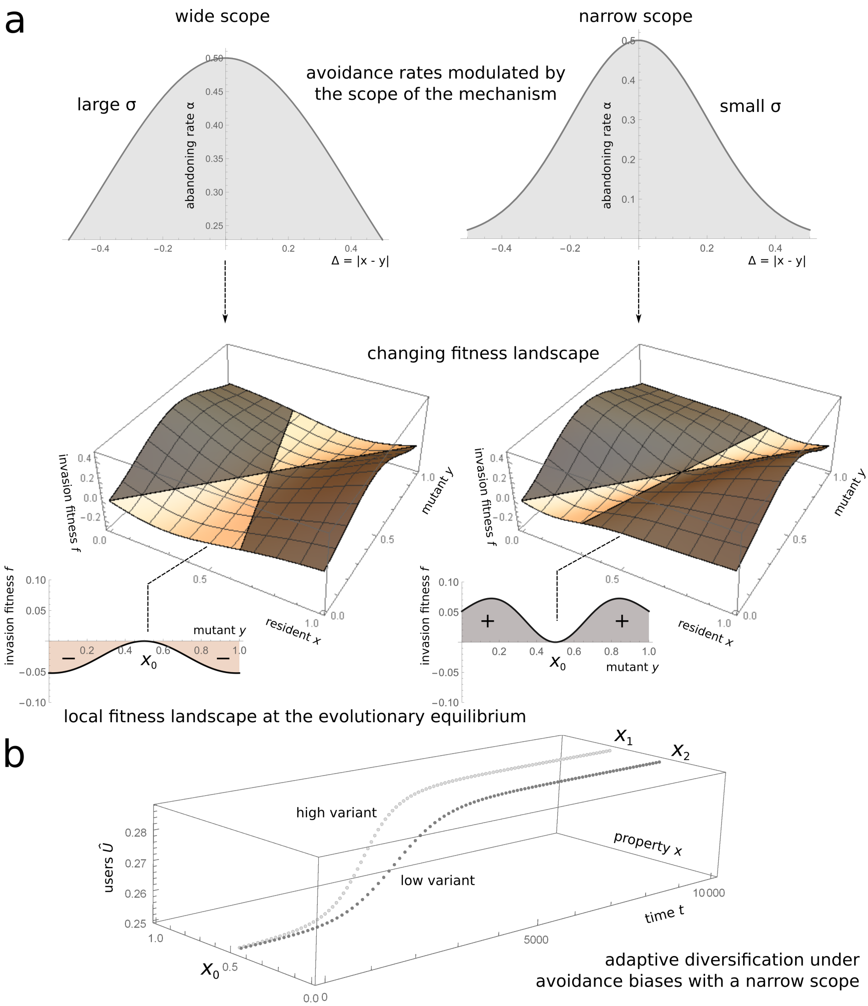


Figure 1. (a) Evolutionary dynamics for wide (left, $\sigma = 0.4$) and narrow (right, $\sigma = 0.2$) scopes. Top panel: avoidance rates $\alpha$ are modulated by the scope of the mechanism. Mid panel: changing fitness landscape showing invasion fitness $f(x, y)$ for all combinations of a predominant variant ('resident') characterized by $x$ and a rare variant ('mutant') characterized by $y$. Dark regions denote positive, and light regions negative invasion fitness. In both cases, $x_0$ is an evolutionary *attractor*. This is because below/above $x_0$, mutants closer to $x_0$ than the resident have positive invasion fitness, so that $x$ is driven towards $x_0$. Bottom panel: Fitness landscapes of the mutant variant $y$ in an environment set by the resident $x_0 = 0.5$ at the evolutionary equilibrium. In the right case, $x_0$ is a minimum of the fitness landscape, can be invaded by nearby variants and is hence a *branching point*. (b) Diversification into two variants if the underlying avoidance mechanism has a narrow scope (small $\sigma$). NB: the symmetry in (b) is a direct consequence of the symmetry of $\lambda(x)$ in the present model which we assume for simplicity (the shape of $\lambda(x)$ can be more complex, of course). The same holds true for the location of $x_0$ which we simply assume to be $0.5$ in this simulation.

same, they will be dropped in order to behave in an extravagant way. But if individuals have a very fine-grained perception of different variants and can tease them apart easily then they will stick to their own behavior because, after all, it is then more likely to be judged as extravagant by others. Thus, fine-grained perception of formal and semantic differences between linguistic constituents which—*given the mechanisms discussed here*—can ensure the diversity of linguistic (phonological, lexical, constructional) inventories.[2]

The model proposed in this paper clearly is simplistic. It builds on homogeneous mixing, simple population structure, and the technical assumption that linguistic properties evolve gradually (in other words, that changes are small and that innovations spread fast). Most importantly, it only explicitly features mechanisms of negative but no mechanisms of positive frequency dependence, respectively, such as conformity biases (on the pragmatic level), priming effects (on the cognitive-psychophysiological level), or simplicity preferences inducing positive frequency dependence. Obviously, such mechanisms exist as well and shape linguistic evolution (Jäger & Rosenbach, 2008; Enfield, 2008; Baumann & Sommerer, 2018). Our point is not that avoidance biases are *necessary* for explaining linguistic diversity. The point is that, everything else being equal, mechanisms of negative frequency dependence can yield interesting evolutionary dynamics in that they represent *sufficient* conditions for linguistic diversification.

## References

Baumann, A., & Sommerer, L. (2018). Linguistic diversification as a long-term effect of asymmetric priming: an adaptive-dynamics approach. *Language Dynamics and Change*, *8*(2), 253–296.

Brisson, D. (2018). Negative frequency-dependent selection is frequently confounding. *Frontiers in Ecology and Evolution*, *6*.

Brugmann, K. (1917). Gleichklangvermeidung in der lautgesetzlichen Entwicklung und in der Wortbildung. *Indogermanische Forschungen (1917)*, *38*(1917), 117–128.

Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual review of Ecology and Systematics*, *31*(1), 343–366.

Dercole, F., & Rinaldi, S. (2008). *Analysis of evolutionary processes.* PUP.

Doebeli, M. (2011). *Adaptive diversification.* PUP.

Doebeli, M., & Ispolatov, I. (2010). A model for the evolutionary diversification of religions. *Journal of theoretical biology*, *267*(4), 676–684.

Enfield, N. (2008). Transmission biases in linguistic epidemiology. *Journal of Language Contact*, *2*(1), 299–310.

---

[2] As a corollary, populations with impeded perceptive sensitivity (for whatever reason, be it environmental, technological, or, thinking of language evolution on the biological time-scale, physiological or cognitive) should accommodate less diverse inventories.

Fajardo, J. A. S. (2019). Exploring the 'shashification'of teenage slang. *English Today*, *35*(3), 49–54.

Geritz, S. A., Metz, J. A., Kisdi, É., & Meszéna, G. (1997). Dynamics of adaptation and evolutionary branching. *Physical Review Letters*, *78*(10), 2024.

Günther, C., Kotowski, S., & Plag, I. (2020). Phrasal compounds can have adjectival heads: evidence from English. *English Language & Linguistics*, *24*(1), 75–95.

Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguistics*, *37*(6), 1043–1068.

Haspelmath, M. (2000). The relevance of extravagance: A reply to Bart Geurts. *Linguistics*.

Hilpert, M., & Saavedra, D. C. (2018). The unidirectionality of semantic changes in grammaticalization: an experimental approach to the asymmetric priming hypothesis. *English Language & Linguistics*, *22*(3), 357–380.

Jäger, G., & Rosenbach, A. (2008). Priming as a testing ground for historical linguists?–A reply to Chang, Eckardt and Traugott. *Theoretical Linguistics*, *34*(2), 143–156.

Meszena, G., Kisdi, E., Dieckmann, U., Geritz, S. A., & Metz, J. A. (2002). Evolutionary optimisation models and matrix games in the unified perspective of adaptive dynamics. *Selection*, *2*(1-2), 193–220.

Page, K. M., & Nowak, M. A. (2002). Unifying evolutionary dynamics. *Journal of Theoretical Biology*, *219*, 93–98.

Petré, P. (2017). The extravagant progressive: an experimental corpus study on the history of emphatic [be ving] 1. *English Language & Linguistics*, *21*(2), 227–250.

Rohdenburg, G. (2011). Cognitive complexity and horror aequi as factors determining the use of interrogative clause linkers in english. In *Determinants of grammatical variation in english* (pp. 205–250). De Gruyter Mouton.

Szmrecsanyi, B. (2008). *Morphosyntactic persistence in spoken English: A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis.* De Gruyter Mouton.

Ungerer, T., & Hartmann, S. (2020). Delineating extravagance: Assessing speakers' perceptions of imaginative constructional patterns. *Belgian Journal of Linguistics*, *34*(1), 345–356.

Vosberg, U. (2003). The role of extractions and horror aequi in the evolution of -ing complements in modern english. In *Determinants of grammatical variation in english* (pp. 305–328). De Gruyter Mouton.

Walter, M. A. (2007). *Repetition avoidance in human language.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Walter, M. A. (2008). Repetition leads to lenition in consonant sequences. *8th International Seminar on Speech Production*, 209–212.

# THE ROLE OF TRUST IN LANGUAGE EMERGENCE

Ted Bayne
Independent Researcher
tedbayne@comcast.net
Concord, MA 01742 USA

Over the past 25 years Chris Knight, Jerome Lewis, James Hurford, and others have proposed community-wide trust and social egalitarianism as the critical conditions for the innovation and enculturation of spoken language (gender political balance is tied into egalitarianism). This paper evaluates the underpinnings and predispositions for these developments. We contend that inter-subjective trust heightens the *efficacy* of cognition in semiotic inference. Human trust has particularly deep evolutionary antecedents in the entanglement of social and personal domains within cultural and biological factors. Millions of years of primate and hominin *mind-reading* (ToM) laid down tracks for an eventual transition to human speech using the same inferential *socio-cognitive* channels (at very different levels).

## 1. Introduction.

The cultural achievement of social trust ultimately leverages bio-genetic evolution preceding and underpinning it but, as *cultural*, it has a complex set of dependencies in which the primary factor seems to be a socio-political egalitarianism (Knight & Lewis, 2017; Boehm, 2009). Moreover, this egalitarian thread has dependencies that beg for explanation. At some point, *H. sapiens* launched into a "culture only" evolutionary phase (Welsch, 2016). There is debate about when this happened. Evidence suggests that it happened slowly but was in place *before* the migration out of Africa (~60 kya; Domínguez-Andrés & Netea, 2019).

The history and field work forming the stimulus for this paper primarily have studied sub-Saharan and south African immediate-return hunter-gatherers.

## 2. A Protracted Process on the way to Trust: Preconditions

The hypothesis of a combination of community-wide trust, social egalitarianism, and gender political balance has been developing a growing appeal. *H. sapiens* shows in early traces the capacity to accumulate culture over generations. Trust eased and supported the *innovation* of shared spoken language. But these times of trust were the last act in a play with a hundred acts: key underpinnings and precursors were needed. As Deacon asserts (2017), no new brain structures evolved to support language or its correlates (Christiansen, 2013). So, from an evolutionary point of view, a *culture-only* evolutionary history was called for to achieve these outcomes but this evolutionary process was highly complex: "a very wide spectrum of entangled conditions is required" (Dor, Knight, & Lewis 2014). This *spectrum* includes social, political, cognitive, emotional as well as cultural factors and there are implicit sub-dependencies between those factors. This paper aims to expose these underpinnings.

## 3. Subcortical & Autonomic Predispositions to Sociality and Proximity

Critical to the later emergence of Trust is a multifaceted set of evolutionary developments, some in primate development and some later in the hominin tree that relate to the toleration of closeness and then an expansion of cooperative closeness into inter-subjectivity.

**3.1. Social engagement's *autonomic* capacities**: an autonomic social-engagement system (Porges & Carter, 2013) developed in the primate line over millions of years. This system developed long after the much earlier parasympathetic and sympathetic autonomic systems. Among primates, bonobos outperform chimps in social intelligence showing this emergent set of capacities. (Krupenye et al, 2017; Gruber et al, 2016). In a provocative example of this, bonobo sexual beckoning gestures combine *deixis* and *iconicity* (Clay & Genty, 2017). Porges summarizes this autonomic system in this passage [key is the *visceral sensing of safety* - required for trust]:

> *This synergism of neural mechanisms in mammals down-regulated defensive systems and promoted proximity by providing social cues (e.g., intonation of vocalization, facial expressivity, posture, and head gesture) that the organism was not in a physiological state that promoted aggressive and dangerous behaviors. Detection of these social cues allowed for symbiotic regulation of behavior and the elaboration of reciprocal care-giving.* (Porges & Carter, 2013)

The role of social hormones (e.g. oxytocin) in this area has often been pushed to center stage. But reducing anxiety about social proximity cannot by itself claim so much of the credit for such a broad and creative spectrum of behavioral and communicative social invention. Human neuroplasticity arising in a social setting *without* a straight-jacket of "fixed behaviors" builds these forms of trust. (a relevant discussion is in Hurford, 2007)

**3.2. Interaction engine** (Levinson, 2006): The main thrust of Levinson's "interaction engine" is that humans engaged in enhanced communication exercises *before* formal language emerged. *H. sapiens* developed this semiotic interactivity on top of the autonomic social engagement capacities mentioned above. Attention to the Other occurred as an early precursor to verbal exchanges (Hrdy, 2011). Elementary mind-reading (ToM) developed in the primate line reinforcing this. (Vogeley, 2017; de Waal, 2022). As an example, Brinck stresses that, from a very early age, infants are poised to respond to communicative engagements, shifting attention and gaze vis-à-vis objects in terms of the caregiver (Brinck et al, 2017). Also, interactive play is a critical modality with deep roots and is linked to prosocial intelligence (Kuczaj et al, 2008).

**3.3. Intersubjectivity**: (de Waal, 2019) The claim of Trust in a theory of language-emergence is congruent with accounts of early ToM and forms of communication that leverage inferential sharing of intent-signals (non-verbal). The cognitive substrates for these faculties are critical. Deep in the primate line are forms of proto-inter-subjectivity in genus *Pan* (Krupenye et al, 2017; Kano et al, 2017). Interpreting the intent of the Other exercises faculties of domain-general cognitive inference that work well within a wide spectrum of communicativeness. What evolved was not *discretely coded* signals "one at a time". What evolved was a multi-purpose meta-representational *capacity* that could work with any signal system because it evolved cognitive strengths that grew increasingly adept at "getting" what the Other intended across multiple modalities. Only humans use *ostensive* communication while nonhuman primates are mainly *intentional* (Scott-Phillips, 2015). With trust, humans cooperate in sharing an inferential *common ground* that can draw on a vast data store. Trust endows human participants with a critical measure of *efficiency* in both innovation and early usage phases of this communications adventure (Knight, 2018).

**3.4. Attachment theory** (Hrdy, 2009) reveals the relational/cognitive strengths of human children that support their ability to securely attach in *alloparental* settings (Kramer et al, 2018; Gopnik, 2016a). The *child's* autonomous contribution to this hugely advantageous evolutionary pathway is often overlooked. In their summary of this subject, the authors assert "These observations clearly support Hrdy's emphasis on the extensiveness and importance of alloparental behavior among hunter-gatherer groups…" (Hewlett & Lamb, 2005). The ToM strengths of the young child (<6 yrs.) (Nonnenmacher et al, 2021) presage the same capacities emerging in adults (Dörrenberg et al, 2018). The child can appraise the safety and emotional signals of the caregiver towards herself. Infant attachment styles later become a psychological advantage for the adult (Fraley, 2019). Trust between adults is affected here as well. Children achieve security with non-maternal caregivers using perceptive faculties that assess the Other's emotional intent.

**3.5. Spoken Language** (evolving later) grew to leverage a semiotic-semantic-metaphoric common ground which scaffolded an early ritual and multi-modal expressiveness – dance/song, oro-facial and body gesture (Lewis, 2014: §7.5; Prieur et al, 2019). This semiotic scaffolding was embedded in the cultural practices that sought to protect egalitarianism and its roots in the economy, reproductive politics, the hunt, and the spirits of the forest. For ages, preverbal inter-subjectivity (ToM) used domain-general inferential faculties (multi-sensory) to grasp communicative intent: taking in contextual cues from multiple channels. Cultural life was already rich in semiotic/semantic meanings.

> *...no new brain structures evolved to support the distinctive cognitive capacities of the human species such as language. Instead, language functions have effectively recruited older neural systems, previously adapted to serve other functions.* (Deacon, 2017)

As *H. sapiens* evolved along a culture-only path, social communication increasingly enjoyed a plasticity – a semiotic continuum (Lewis in Arbib, 2013; Malafouris, 2016: §4). Modalities other than verbal speech can be ostensive and have communicative intent. The pre-verbal inferential faculty is congruent with and segues into the later verbal ostensive-inferential faculty (Scott-Phillips, 2015). The sound-stream of words becomes yet another modality to blend with available (usual or new) cues and scaffolding. The “digital” aspects of verbal speech don’t deny what is called the “invariance problem”. The interpretation of this sound-stream relies on the cues and contexts that disambiguate the stimulus – to leverage this scaffolding is essential.

Jerome Lewis in his study of the Mbendjele goes right to the heart of this semiotic plasticity and the prominence of their multi-faceted musicality:

> *“Mbendjele have developed specific styles of communication for different audiences and situations. They mix words with sung sounds, ideophones, expletives, whistles, signs, hand signals, gestures, vocabulary from other people’s languages, animal sounds, and other environmental sounds, sometimes in a single speech act. In the context of forest hunting and gathering, the role of different language styles and communicative strategies suggests that diverse styles of communicating could have been crucial to the survival of early humans...”* (Lewis in Botha & Knight, 2009)
>
> *[For the Mbenjele] ...it is not what people are singing but the polyphonic yodelling singing style, not which dances they dance or which spirits they call but the ritual structures they follow, not the language they speak but how it is spoken. The perception of what it means to be Yaka is based on an aesthetic quality and lifestyle as much as on genealogical accident, a distinctive sense of style in which music is more central to culture than language.* (Lewis, 2002)

## 4. Relaxed Selection and Self-Domestication

Important protracted developments resulted in less over-determination of *Homo* by hardwired genes and more by environmental responses coming from neuro-plasticity and epigenetics (Deacon, 2017). "Self-domestication" and neoteny (de Waal, 2022; Bednarik, 2011) form two major themes in this long term process (Hare, 2017). This evolutionary process goes back to forms of our genus roughly 2 mya. Deacon places "relaxed selection" in that time frame as well as growth-spurts in stone tool use and inter-continental migration (Deacon, 2016).

> *Although traditional assumptions about the role of genetic change in evolution have tended to focus on mutations that augment some function, evidence is growing that gene duplication and gene loss – including especially loss of non-coding regulatory sequences – has contributed to significant evolutionary change… Human-specific loss of over 500 otherwise highly conserved non-coding regulatory sequences has recently been reported (McLean et al 2011). This extensive loss of genetic regulation may be a signal of human-specific relaxation of selection and an increased sensitivity to epigenetic and environmental influences."*
> (Deacon, 2016)

The net effect of self-domestication is the release of the organism from pre-set genetic constraints or fixed behavioral plans (Gómez-Robles et al, 2015). There are connections between the license to innovate, pretend-play, and ritually sing & dance just as examples (Hare, 2017). The relaxation of genetic constraints means that Play can bridge to multiple cultural, political, and interpersonal domains in multi-modal ways. For instance, rather than have an aggressive single-minded rigidity about issues of sexual politics, humorful imagination can enter into the available responses and bring the female cohort together making reverse dominance an exhausting **and** exhilarating exercise.

## 5. Plasticity, Play, Egalitarianism, and Trust

Trust is tacitly built into play. Playmates "play along" honoring unspoken rules. If the rules get extended, playmates might pick up on this extension and accommodate the new "twist". Play draws on a prior and very long development of cooperation in the genus *Homo* (Tomasello, 2014) along with a reduction of both the proactive and reactive types of aggression in *H. sapiens* (Wrangham, 2018).

Particularly among immediate-return hunter-gatherers, the ethos of sharing and cooperating is very strong. There is an implicit element of trust in the cooperation in an immediate-return economy. The presence of trust in Play would be as palpable. Baka children's play exhibits values like egalitarianism that are strongly embraced in the adult culture (Kamei, 2005).

Play in children and adults is an opportunity stemming from the loss of genetic regulation: invented vocalizations versus fixed calls, for instance. While the self-domestication and relaxed selection mentioned earlier took place significantly before the period of culture-only or proto-culture-only evolution (Welsch, 2016), play in child-development or adult ritual would need cultural supports to become stable or habitual parts of human life.

> *"... in the safe and well-provisioned context of play, characteristic of a species with an extended juvenile period, individuals place themselves into unconventional and often disorienting positions or orientations. These novel situations afford opportunities to experiment with a variety of behavioral and cognitive routines and generate novel, and possibly adaptive, responses, or modules."* (Pellegrini & Pellegrini, 2013)

Play draws on intuitive and creative cognitive and affective resources. It perpetuates an egalitarian setting and asserts it anew.

"Social play – play involving two or more playmates – is necessarily egalitarian. It always requires a suspension of aggression and dominance and heightened sensitivity to the needs and desires of the other individual involved." (Gray, 2013)

Closely related to the suspension of aggression is the matter of self-control. "[The] same cultures that allowed their children the greatest freedom to play also produced people who seemed to have the greatest capacity for self-control. … [The] … amounts of play engaged in by hunter-gatherer children helped to promote their extraordinary capacities to regulate their emotions in ways that were adaptive to their conditions of life." (Gray, 2013)

## 5.1. Play Liberates Ritual Content

Rituals still practiced in extant groups (Congo basin) are created with an imagination driven by Play. Shared aesthetic rules allow autonomy and humorful coordination. Play allows the moral themes of reverse dominance and female solidarity to be explicitly danced and sung, together with impersonations that bring on wild laughter. Play allows performance to be a virtual space where creativity is free and enthusiastic. Even an *origins* story is playful and whimsical yet evocative of "real life". Named rituals (e.g. for the Mbendjele: "Ngoku" and their origins story) are embodied metaphorical figures (without prescribed lyrics) that evoke the moral and spiritual narratives lived by the band.

## 6. Female Strength and Role in the Hunt

Women are not excluded. They are often active in the hunt and then they are the "arms" of the nets when they are thrown to catch game. They have a spiritual-telepathic channel to the creatures. This channel is why the very antipathy between menstrual blood and the blood of the hunted represents the potency of their communicative relationship with those creatures (Finnegan, 2016). *"it is overwhelmingly women who locate forest animals. It is through a privileged relationship with spirits that women participate in the hunt."* (Finnegan, 2016)

When women perform the *bobanda* ritual to recover the men's hunting luck or reset social homeostasis in the community, they are revered for their strength. Their role in these works is a positive and assertive one.

Cooperative childcare combined with women's rightful claim to a share of the meat forms the sexual core of the egalitarianism kept alive by ritual vigilance. Evidence shows that in early kinship systems, childbearing women mostly raised their young collectively in a matrilocal setting (Knight, 2011; Hawkes et al, 1998) and were better poised there to obtain services from interested males. Climatic adversity and population pressures often increased the incentive to "go home" (Kim et al, 2014). Vigilance of the female cohort began with an early campaign by women to get control over sexual matters. The history and timeline of bodypainting is catalogued and analyzed thoroughly in Watts' "The Red Thread" (2014). The sex-strike boldly dissimulating the sexual status quo is a "moral strategy" (Finnegan, 2016).

The species of trust that is possible at this level of political counterpoint does not emerge from sentimentalism. It emerges from women's playful confidence and self-esteem as well as their sense of reproductive potency in the community. They know well the virulent danger of the alternative: they're always on the lookout.

Pretend-play is an alternative to alpha-male power with its competition and fear. Importantly, Play in the Ngoku ritual dance-song shows an affirmation of plasticity, possible when personal autonomy is embraced along with its culture (Lewis, 2013). Given the female cohort's continual vigilance coupled with its cooperative bonds (Finnegan, 2016), the resulting egalitarianism yields an ecology of safety producing the "good sounds" that keep the people and the forest in peace. (Lewis, 2014; Pellegrini & Pellegrini, 2013). This safety is needed for language innovation.

## 7. Overview

The several threads discussed above are mutually tied up in the emergence of a functioning human trust. Certain long term overarching developments were essential: relaxed selection, self-domestication, neoteny, autonomic support for social engagement (hormonal and neurological), plasticity and epigenetics (vs.

genetically “hardwired”). The growth and inheritance of cultural forms became a follow-on protracted “evolution” with its own genetic consequences. Ultimately an evolving *culture-only* social process took center stage. Socio-cultural legacies underpinning trust included gender balance and egalitarianism, play, attachment, forms of inter-subjectivity as well as cooperation in multiple domains. Trust opens up human cognitive pores to a broad range of commonly held realities and fantasies. Inferential grasp of the Other’s *gist* is emboldened by trust (Gweon, 2021). Proto-language (and beyond “proto”) draws on an intimate and patient sharing of ostensive behaviors and resulting inferences – all having deep roots in the history of ToM in our genus. “In an evolving hominin species, language will not even begin to evolve unless and until intensified levels of community-wide trust and a shared virtual domain have been put in place.” (Knight & Lewis, 2017)

## References

Allen, N. J., Callan, H., Dunbar, R., James, W. (eds.) (2011). *Early human kinship: From sex to social reproduction*. John Wiley & Sons.

Arbib, M.A. ed. (2013). *Language, music, and the brain: A mysterious relationship.* Cambridge, MA: MIT Press.

Bednarik, R. G. (2011). *The human condition*. Springer Science & Business Media. (see pp. 159-64, Section *Ontogenic Plasticity and Neoteny*)

Boehm, C. (2009). *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press.

Botha, R. & Knight, C. (2009). *The cradle of language*. Oxford: Oxford University Press.

Botha, R., & Everaert, M. (eds.). (2013). *The evolutionary emergence of language: Evidence and inference* (Vol. 17). Oxford Studies in the Evolution of Language.

Brinck, I., Reddy, V., & Zahavi, D. (2017). The primacy of the “we”. In Durt et al, 2017 (ch. 6).

Christiansen, M. H. (2013). Language has evolved to depend on multiple-cue integration. In Botha et al (eds.) (2013). *The evolutionary emergence of language: Evidence and inference.* pp. 42-61.

Clay, Z., & Genty, E. (2017). Natural communication in bonobos: Insights into social awareness and the evolution of language. In Hare, B. & Yamamoto, S. (eds.) (2017) ch. 8.

Deacon, T. W. (2009). Relaxed selection and the role of epigenesis in the evolution of language. In Blumberg, M., Freeman, J., & Robinson, S. (eds.) 730-752. *Oxford handbook of developmental behavioral neuroscience*. New York: Oxford University Press.

Deacon, T. W. (2010). A role for relaxed selection in the evolution of the language capacity. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 9000-9006.

Deacon, T. W. (2016). On human (symbolic) nature: How the word became flesh. In Etzelmüller, G., & Tewes, C. (eds.). (2016:129-50).

Deacon, T. W. (2017). Human variability and the origins and evolution of language. In Tibayrenc et al (eds.) (2017) *On Human Nature…* p. 557.

Domínguez-Andrés, J., & Netea, M. G. (2019). Impact of historic migrations and evolutionary processes on human immunity. *Trends in Immunology*, *40*(12), 1105-1119.

Dor, D., Knight, C. & Lewis, J. (2014). *The social origins of language*. Oxford: Oxford University Press.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, *46*, 12-30.

Dunbar, R., Knight, C., & Power, C. (eds.). (1999). *The evolution of culture: An interdisciplinary view*. Rutgers University Press.

Durt, C., Fuchs, T. & Tewes, C. (eds.) (2017). *Embodiment, enaction, and culture: Investigating the constitution of the shared world*. MIT Press.

Enfield, N.J. & Levinson, S.C. (eds.) (2006). *Roots of human sociality*. Oxford: Berg.

Etzelmüller, G., & Tewes, C. (eds.). (2016). *Embodiment in evolution and culture*. Mohr Siebeck.

Etzelmüller, G. (2016a). The lived body as the tipping point between an evolutionary and a historical anthropology. In Etzelmüller & Tewes (eds.). 2016: 205.

Finnegan, M. (2016). Who sees the elephant? Sexual egalitarianism in social anthropology's room. In Power, C., Finnegan, M., & Callan, H. (eds.), 2016: ch.5.

Fraley, R. C. (2019). Attachment in adulthood: Recent developments, emerging debates, and future directions. *Annual review of psychology*, *70*, 401-422.

Gómez-Robles, A., Hopkins, W. D., Schapiro, S. J., & Sherwood, C. C. (2015). Relaxed genetic control of cortical organization in human brains compared with chimpanzees. *Proceedings of the National Academy of Sciences*, *112*(48), 14799-14804.

Gopnik, A. (2016a). *Perspectives on Parenting* (video) (Roots of Empathy Series) https://www.youtube.com/watch?v=_PEB69Ing4w
The subject of "alloparenting" comes up at 12:10 in the video.

Gray, P. (2013). The value of play-filled childhood in development of the hunter-gatherer individual. In Narvaez et al, 2013: ch.14.

Gruber, T., & Clay, Z. (2016). A comparison between bonobos and chimpanzees: a review and update. *Evolutionary Anthropology: Issues, News, and Reviews*, 25(5), 239-252.

Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in Cognitive Sciences*, *25*(10), 896-910.

Hare, B., Wobber, V., & Wrangham, R. (2012). The self-domestication hypothesis: Evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour*, *83*(3), 573-585.

Hare, B. (2017). Survival of the friendliest: *Homo sapiens* evolved via selection for prosociality. *Annual review of psychology*, *68*, 155-186.

Hare, B. & Yamamoto, S. (eds.) (2017). *Bonobos: Unique in mind, brain, and behavior.* Oxford University Press. chaps. 2,3,4 on female dominance.

Hawkes, K., O'Connell, J. F., Jones, N. B., Alvarez, H., & Charnov, E. L. (1998). Grandmothering, menopause, and the evolution of human life histories. *Proceedings of the National Academy of Sciences*, *95*(3), 1336-1339.

Hewlett, B. S., & Lamb, M. E. (eds.) (2005). *Hunter-gatherer childhoods: Evolutionary, developmental, and cultural perspectives*. New Brunswick: Transaction Publishers.

Hewlett, B. S. (2016). Social learning and innovation in hunter-gatherers. In Terashima, H., & Hewlett, B. S. (2016:1-15).

Hrdy, S. B. (2009). *Mothers and others*. Cambridge: Belknap.

Hurford, J. (2007). *The origins of meaning.* Oxford: Oxford University Press.

Kamei, N. (2005). Play among baka children in cameroon. In Hewlett & Lamb (2005:ch.16) pp. 343-59.

Kano, F., Krupenye, C., Hirata, S., Call, J., & Tomasello, M. (2017). Submentalizing cannot explain belief-based action anticipation in apes. *Trends in cognitive sciences*, *21*(9), 633-634.

Kim, H. L., Ratan, A., Perry, G. H., Montenegro, A., Miller, W., & Schuster, S. C. (2014). Khoisan hunter-gatherers have been the largest population throughout most of modern-human demographic history. *Nature communications*, *5*, 5692.

Knight, C. (2011). Early human kinship was matrilineal. In Allen et al (2011:61-82).

Knight, C. (2014). Language and symbolic culture: an outcome of hunter-gatherer reverse dominance. In Dor, D., Knight, C. & Lewis, J. (2014: 228-46).

Knight, C. (2018). Pressure for trust-based efficiency shaped the evolution of language. In *The Evolution of Language: Proceedings of the 12th International Conference (EVOLANGXII).*

Knight, C., & Lewis, J. D. (2017). Wild voices: Mimicry, reversal, metaphor, and the emergence of language. *Current Anthropology*, *58*(4) pp. 435-53.

Kramer, K. L., & Veile, A. (2018). Infant allocare in traditional societies. *Physiology & behavior*, *193*, 117-126.

Krupenye, C., MacLean, E.L. & Hare, B. (2017). Does the bonobo have a (chimpanzee-like) theory of mind? In Hare & Yamamoto (2017). Ch. 6.

Levinson, S.C. (2006). On the human "Interaction Engine". In Enfield & Levinson, 2006:39-69.

Lewis, J. (2008). Ekila: blood, bodies, and egalitarian societies. *Journal of the Royal Anthropological Institute*, *14*(2), 297-315.

Lewis, J. (2013). A cross-cultural perspective on the significance of music and dance to culture and society: Insight from BaYaka Pygmies. In Arbib, M.A. ed. (2013:45-67).

Lewis, J. (2014). BaYaka Pygmy multi-modal and mimetic communication traditions. A cross-cultural perspective on the significance of music and dance to culture and society: Insight from bayaka pygmies. In Dor, D., Knight, C. & Lewis, J. (2014: 77-91).

Malafouris, L. (2016). On human becoming and incompleteness: a material engagement approach to the study of embodiment in evolution and culture. In Etzelmüller & Tewes, 2016:289-305.

McLean, C. Y., Reno, P. L., Pollen, A. A., Bassan, A. I., Capellini, T. D., Guenther, C., ... & Wenger, A. M. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature*, *471*(7337): 216-9.

Narvaez, D. et al (eds.) (2013). *Evolution, early experience and human development*. Oxford: Oxford University Press.

Nonnenmacher, N., Müller, M., Taczkowski, J., Zietlow, A. L., Sodian, B., & Reck, C. (2021). Theory of Mind in Pre-school Aged Children: Influence of Maternal Depression and Infants' Self-Comforting Behavior. *Frontiers in psychology*, *12*.

Pellegrini, A.D. & Pellegrini, A.F.A. (2013). Play, plasticity, and ontogeny in childhood. In Narvaez et al, 2013: ch.13.

Porges, S.W. & Carter, S. (2013). In Narvaez et al, 2013: ch.5.

Power, C., Finnegan, M., & Callan, H. (eds.). (2016). *Human origins: contributions from social anthropology* (Vol. 30). Berghahn Books.

Prieur, J., Lemasson, A., Barbu, S., & Blois-Heulin, C. (2019). History, development and current advances concerning the evolutionary roots of human right-handedness and language: Brain lateralisation and manual laterality in non-human primates. *Ethology*, *125*(1), 1-28.

Scott-Phillips, T. (2015). *Speaking our minds*. New York: Palgrave MacMillan.

Terashima, H., & Hewlett, B. S. (2016). *Social learning and innovation in contemporary hunter-gatherers*. Tokyo: Springer.

Tibayrenc, N. & Francisco J. Ayala F.J. (eds.) (2016). *On human nature: Biology, psychology, ethics, politics, and religion.* Academic Press.

Tomasello, M. (2014). The ultra-social animal. *European journal of social psychology*, *44*(3), 187-194.

Vogeley, K. (2017). Two social brains: neural mechanisms of intersubjectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1727), 20160245.

Watts, I. (2014). The red thread: pigment use and the evolution of collective ritual. In Dor, D., Knight, C. & Lewis, J. (2014: 225-6).

de Waal, F. B. (2019). *Mama's Last Hug: Animal emotions and what they tell us about ourselves*. New York: Norton.

de Waal, F. B. (2022). *Different: Gender through the eyes of a primatologist.* New York: Norton.

Welsch, W. (2016). Bodily changes during the proto-cultural period and their ongoing impact on culture. In Etzelmüller & Tewes, 2016: 353-63.

Wrangham, R. W. (2018). Two types of aggression in human evolution. *Proceedings of the National Academy of Sciences*, *115*(2), 245-253.

# EARLY 'LANGUAGE HOMOLOG' BRAIN ASYMMETRY IN NEWBORN MONKEYS PAVES THE WAY FOR THE DEVELOPMENT OF GESTURAL COMMUNICATION'S LATERALIZATION

YANNICK BECKER*[1], ROMANE PHELIPON[2], SIHAM BOUZIANE[1], JULIEN SEIN[3] , LIONEL VELLY[3], LUC RENAUD[3], EMILIE RAPHA[3], ADRIEN MEGUERDITCHIAN[1,4]
*Corresponding Author: yannick.becker@gmx.net

[1]Laboratoire de Psychologie Cognitive, Aix-Marseille Univ, CNRS UMR 7290, Marseille, France
[2]Centre de Recherche en Neurosciences de Lyon, Université Lyon 1, INSERM U1028 CNRS UMR5292, Lyon, France
[3]Institut de Neurosciences de la Timone, Aix-Marseille Univ, CNRS UMR 7289, Marseille, France
[1,4]UAR846 Station de Primatologie CNRS, Rousset, France

The Planum Temporale (*PT*) is a key language area and is structurally and functionally left-lateralized in pre-linguistic infants (Dehaene-Lambertz et al., 2002; Dubois et al., 2010). Therefore, the *PT* was considered a marker of the uniquely human pre-wired brain for language acquisition. This latter hypothesis has been questioned by recent studies reporting similar structural asymmetries in newborn baboons (Becker et al., 2021; 2022). Nevertheless, its functional implication in monkeys, and their potential shared properties with any language features related to *PT* asymmetry in humans remains unknown.

In the present followed-up study in the same baboons (*Papio anubis)* living in social groups at the Station de Primatologie CNRS, we found a clear link between this early brain feature and communicative gesturing. In fact, we show that direction of early PT grey matter asymmetry in newborn baboons predicts the development of later communicative gesture's manual lateralization once juveniles (but not of handedness for non-communicative manipulative action).

Specifically, we investigated intra-individual brain/behaviour correlates within 26 healthy infant baboons (10 females, 16 males). Behavioral measurements included both handedness for communicative gestures *versus* handedness for non-communicative bimanual object manipulation from data collected from 9 months to 3 years old. Brain measurements included previously published MRI data on *PT* grey matter volume interhemispheric asymmetry at the earliest “newborn" age class (i.e., 4 days to 2 months old before the full maturation of myelin, synapses and cell bodies + an outlier of 165 days of age, see Becker et al., 2021 for the full MRI acquisitions and region delineation procedures). Handedness for manipulative actions was assessed using the bimanual coordinated “Tube task” (Hopkins et al., 1996). Communicative gesture was defined as a movement of the hand directed to a specific partner or audience in order to affect its behavior (Molesti et al., 2019). The “Hand slapping” gesture was previously found optimal for measuring such gestural communication’s lateralization in the baboon (Meguerditchian et al., 2013). For each subject and both behaviors, a handedness index of the left (L) and the right (R) hand was computed HI = (R – L) / (R + L) with the sign indicating the direction of asymmetry (negative: left side, positive: right side) and the value, the strength of asymmetry. Subjects were classified into two groups, according to direction of *PT* asymmetry at birth (19 typical left-lateralized, 8 atypical non- or right-lateralized). One sample t-test highlighted a significant positive handedness for communication for the typical *PT* biased group ($p < .006$), which was not found for the atypical *PT* biased group ($p > .3$). Two-sided t-test demonstrated a significant difference between groups ($p < .042$). No effect was found for non-communicative actions ($p > .5$). Logistic regression significantly modeled the effect of the asymmetry quotient of the Planum Temporale at birth on the likelihood of becoming left or right-handed in communicative gesture ($p < .028$). In other words, newborn monkeys with an early larger left-than-right *PT* were more likely to developed right-handed gestural communication once juvenile.

This finding suggests that early *PT* asymmetry in both monkeys and humans might be a pre-wired brain marker of a shared functional feature between human language and monkey’s gestural communication.

## Acknowledgements

The project has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program grant agreement No 716931 - GESTIMAGE - ERC-2016-STG, as well as from grants ANR-16-CONV- 0002 (ILCB), ANR-17-EURE-0029 (NeuroSchool) and the Excellence Initiative of Aix-Marseille University via A*Midex funding (AMX-19- IET-004).

## References

Becker, Y., Sein, J., Velly, L., Giacomino, L., Renaud, L., Lacoste, R., Anton, J.-L., Nazarian, B., Berne, C., & Meguerditchian, A. (2021). Early Left-Planum Temporale Asymmetry in Newborn Monkeys (Papio anubis): A longitudinal structural MRI study at two stages of development. *NeuroImage, 117575.*

Becker, Y., Phelipon, R., Sein, J., Velly, L., Renaud, L., & Meguerditchian, A. (2022). Planum temporale grey matter volume asymmetries in newborn monkeys (Papio anubis). *Brain Structure and Function.*

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional Neuroimaging of Speech Perception in Infants. *Science, 298(5600)*, 2013–2015.

Dubois, J., Benders, M., Lazeyras, F., Borradori-Tolsa, C., Leuchter, R. H.-V., Mangin, J. F., & Hüppi, P. S. (2010). Structural asymmetries of perisylvian regions in the preterm newborn. *NeuroImage, 52(1),* 32–42.

Hopkins, W. D. (1995). Hand preferences for a coordinated bimanual task in 110 chimpanzees (Pan troglodytes): cross-sectional analysis. *Journal of Comparative Psychology, 109(3)*, 291.

Meguerditchian, A., Vauclair, J., & Hopkins, W. D. (2013). On the origins of human handedness and language: A comparative review of hand preferences for bimanual coordinated actions and gestural communication in nonhuman primates. *Developmental Psychobiology, 55(6)*, 637–650.

Molesti, S., Meguerditchian, A., & Bourjade, M. (2019). Gestural communication in olive baboons (Papio anubis): Repertoire and intentionality. *Animal Cognition.*

# VISUALIZATION OF VOCAL FLEXIBILITY IN ASIAN ELEPHANTS

Veronika C. Beeck *[1], Gunnar Heilmann [2], and Angela S. Stoeger [1]

*Corresponding Author: veronika.beeck@univie.ac.at
[1]Department of Behavioural & Cognitive Biology, University of Vienna, Vienna, Austria
[2]MeCalc Technologies GmbH, Berlin, Germany

## Background

Elephants demonstrate both biomechanical and cognitive flexibility in sound production. They are even among the few mammals capable of vocal production learning (Poole et al., 2005; Stoeger et al., 2012). Functions and mechanisms of this flexibility remain largely unknown. The small vocal repertoire of 8-13 call types spans from infrasonic rumbles (F0 8-35 Hz) to high pitched trumpets (F0 300-600 Hz) and, in Asian elephants, species-specific squeaks (F0 300-2300 Hz) (Stoeger & de Silva, 2014). In this project, we studied sound production and how it relates to potential information coding and social learning in Asian elephants, which have been studied much less than African elephants.

## Methods

We used an acoustic camera (gfai tech, Berlin) to visualize sound emission in addition to audio and video data to analyse acoustic structures and body movements in captive Asian elephants. The acoustic camera uses an array of microphones to localize sound sources based on a sum- and delay-beamforming algorithm (Stoeger et al., 2012) and depicts sound pressure levels (SPL) by colour coding (Figure 1).

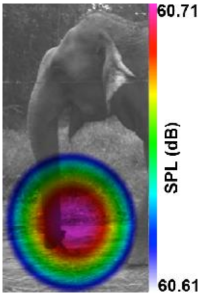


*Figure 1.* Nasal rumble.

## Results

The acoustic camera showed that squeaks were emitted orally in three females ($N_{calls} = 90$) and nasally in one ($N_{calls} = 14$). Our data suggest oral squeaks are produced by vibration of the tensely closed lips, a mechanism unique in the animal kingdom except for humans (Beeck et al., 2021). We recorded squeaks only from 19 out of 56 subjects. One female squeaked through her narrowed nostril, revealing productive flexibility even within one call type. Together, this indicates the involvement of learning in squeak production. Trumpets (uttered through the trunk) and squeaks occurred in contexts of arousal and encoded individual identity. Rumbles in 9 females ($N_{calls}$= 203) were emitted either through the mouth or trunk (Figure 1) or both simultaneously, demonstrating velo-pharyngeal coupling and complex vocal tract resonances that increase the parameter space for potential information coding (Beeck et al., submitted).

## Conclusion

Our results demonstrate how Asian elephants increase acoustic variability through anatomical and cognitive adaptations that overcome morpho-mechanical limitations of laryngeal sound production, where across taxa body size and the related vocal apparatus dimensions determine source and filter parameters. This supports the hypothesis that vocal complexity can be expressed in graded repertoires rather than only repertoire size. We set an important framework for elephants to be included in comparative analyses of the multiple dimensions of vocal complexity and their evolution across socially dynamic species, which include flexibility in sound production, vocal learning, and information coding.

## Acknowledgements

This research was funded by the Austrian Science Fund (FWF) project number W1262-B29, Angela S. Stoeger project number P31034-B29, and a VDS CoBeNe final fellowship to Veronika C. Beeck.

## References

Beeck, V. C., Heilmann, G., Kerscher, M., & Stoeger, A. S. (2021). A novel theory of Asian elephant high-frequency squeak production. *BMC biology, 19*(1), 121. doi:10.1186/s12915-021-01026-z

Beeck, V. C., Heilmann, G., Kerscher, M., & Stoeger, A. S. (submitted). Sound imaging demontsrates velopharyngeal coupling and complex spectral variability in Asian elephants.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature*, *434*(7032), 455-456.

Stoeger, A. S., & de Silva, S. (2014). African and Asian Elephant Vocal Communication: A Cross-Species Comparison. In *Biocommunication of Animals* (pp. 21-39).

Stoeger, A. S., Heilmann, G., Zeppelzauer, M., Ganswindt, A., Hensman, S., & Charlton, B. D. (2012). Visualizing sound emission of elephant vocalizations: evidence for two rumble production types. *PLoS One, 7*(11), e48907. doi:10.1371/journal.pone.0048907

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology, 22*(22), 2144-2148.

# TIMBRE IN MUSIC AND PROSODY: COMPARATIVE AND EVOLUTIONARY CONSIDERATIONS

OLIVER TAB BELLMANN[*1]

[*]Corresponding Author: obellman@smail.uni-koeln.de
[1]Systematic Musicology, Institute of Musicology, University of Cologne, Germany

It has been proposed that music and language share a common prosodic origin (e.g., Brown, 2017). Music and language have different components linked to specific acoustical properties, each of which can be associated with evolutionary functions. If these components are present in both domains as well as non-human vocal communication and share functions, they are likely to be inherited from phylogenetically older communication systems (Masataka, 2009).

In the current paper, I study the hypothesis on a common prosodic origin of music and language by focusing on the domain of timbre through a literature review. As shared components of prosody and music, so far, rhythm and pitch gained much attention in comparative language and music research (e.g., Temperley, 2022; Scharinger & Wiese, forthcoming). However, timbre, another component of prosody, received far less attention in comparative research.

Timbre has para-linguistic functions and plays an important role in speech prosody (Thompson et al., 2012). Its functions include encoding of emotion, based on, e.g., spectral energy distribution and signal-to-noise ratio (Briefer, 2012), and information about the signalers' identity based on individual characteristics of the signalers' formant structure (Taylor & Reby, 2010). Those two functions are shared with those of non-human vocal communication (Taylor & Reby, 2010).

In vocally produced – sung – music, too, timbral cues encode emotion and identity (Scherer, 1995; Erickson, 2018). Communication of emotion through timbre in instrumental music was reported in several studies (e.g., Paquette et al., 2018; Bowman & Yamauchi, 2016). Communication of increased emotional arousal and emotional intensity relies on spectral distribution (Eerola et al., 2012; Wu et al., 2014). Encoding of emotion in instrumental music has been related to perceptual features associated with timbre such as "brightness" and "softness" (Eerola & Vuoskoski, 2012; Juslin & Laukka, 2004). Relevant spectral features for this, such as Spectral Centroid and HF-LF Ratio, are also important acoustical features for encoding arousal in emotional prosody and mammalian vocal communication (Briefer, 2012).

Timbre also plays a role in the identification of music and allows the identification of instrumental music through the individual timbral properties of musical stimuli on various levels - of genres (Casey et al., 2014), whole pieces (Schellenberg et al., 1999), excerpts of pieces (Poulin-Charronnat et al., 2004), melodies (e.g., Schellenberg & Habashi, 2015; Peretz et al., 1998) and single musical sounds (e.g., Suied et al., 2014). Thus, identity is communicated in instrumental music through the individual timbral properties of the music. This is similar to the communication of signalers' identity through individual timbral characteristics seen in prosody and mammalian vocal communication (Handel, 1995; Patil et al., 2012; Taylor & Reby, 2010).

In summary, timbre conveys emotional and identity information in speech prosody, and vocal and instrumental music through the same acoustic cues as in mammalian vocal communication. Thus, I claim that music and language share a common prosodic origin, with communication of identity and emotion as a preserved function found today in para-linguistic aspects of prosody and in music.

## Acknowledgements

I would like to thank Dr. Rie Asano for her helpful comments on earlier versions of this abstract.

## References

Bowman, C., & Yamauchi, T. (2016). Perceiving categorical emotion in sound: The role of timbre. *Psychomusicology: Music, mind, and brain*, *26*(1), 15-25.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1-20.

Brown, S. (2017). A joint prosodic origin of language and music. *Frontiers in psychology, 8,* 1894.

Casey, M., Thompson, J., Kang, O., Raizada, R., & Wheatley, T. (2012). Population Codes Representing Musical Timbre for High-Level fMRI Categorization of Music Genres. In G. Langs, I. Rish, M. Grosse-Wentrup, and B. Murphy (Eds.), *Machine Learning and Interpretation in Neuroimaging* (pp.34-41). Berlin, Heidelberg: Springer

Eerola, T., & Vuoskoski, J. K. (2012). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal*, *30*(3), 307-340.

Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception: An Interdisciplinary Journal, 30*(1), 49-70.

Erickson, M. L., Burchard, J., & Phillips, P. (2018). Can listeners hear who is singing? The development of voice category perception. *Journal of Voice, 32*(4), 459-465.

Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Hearing, 2* (pp. 425-461). San Diego: Academic.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research*, *33*(3), 217-238.

Masataka, N. (2009). The origins of language and the evolution of music: A comparative perspective. *Physics of Life Reviews*, *6*(1), 11-22.

Paquette, S., Takerkart, S., Saget, S., Peretz, I., & Belin, P. (2018). Cross-classification of musical and vocal emotions in the auditory cortex. *Annals of the New York Academy of Sciences*, *1423*(1), 329-337.

Peretz, I., Gaudreau, D., & Bonnel, A. M. (1998). Exposure effects on music preference and recognition. *Memory & cognition, 26*(5), 884-902.

Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurell, F., Vieillard, S., & McAdams, S. (2004). Effects of a change in instrumentation on the recognition of musical materials. *Music Perception, 22*(2), 239-263.

Scharinger, M. & Wiese, R. (forthcoming). *How Language Speaks to Music: Prosody from a Cross-domain Perspective*. De Gruyter.

Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory & Cognition, 43*(7), 1021-1031.

Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic bulletin & review, 6*(4), 641-646.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of voice, 9*(3), 235-248.

Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology, 280*(3), 221-236.

Temperley, D. (2022). Music and Language. *Annual Review of Linguistics, 8,* 153-170.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, *109*(46), 19027-19032.

Wu, B., Horner, A., & Lee, C. (2014). The correspondence of music emotion and timbre in sustained musical instrument sounds. *Journal of the Audio Engineering Society*, *62*(10), 663-675.

# COGNITIVE AND GENETIC CORRELATES OF A SINGLE MACRO-PARAMETER OF CROSSLINGUISTIC VARIATION

ANTONIO BENÍTEZ-BURRACO*[1], CANDY CAHUANA[1], SIHAN CHEN[2],
DAVID GIL[3], LJILJANA PROGOVAC[4], JANA REIFEGERSTE[5]
and TATIANA TATARINOVA[6]

*Corresponding Author: abenitez8@us.es
[1]Department of Spanish, Linguistics & Theory of Literature, University of Seville, Spain
[2]Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, USA
[3]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[4]Linguistics Program, Wayne State University, Detroit, USA
[5]Potsdam Research Institute for Multilingualism, University of Potsdam, Germany, and Brain and Language Lab, Georgetown University, Washington, D.C., USA
[6]Department of Biology and Computational Biology, University of LaVerne, USA

Evidence exists for a correlation, and perhaps also causation, between specific linguistic and societal features, in particular those relating to *exoteric* (open) vs. *esoteric* (close-knit) society types, characterizable in terms of population size, mobility, communication across distances, etc. Broadly speaking, languages associated with exoteric societies, or *Type A languages*, have been reported to exhibit less complex phonologies and morphologies, but more complex and more layered syntaxes, with more specialized and obligatory grammaticalized distinctions, while languages associated with esoteric societies, or *Type B languages*, exhibit a complementary clustering of features, including simpler and less layered syntaxes, but more complex phonologies and morphologies, with more irregularity, and more formulaic/memorized language chunks (e.g. Wray & Grace, 2007). We conducted an exhaustive quantitative analysis drawing upon *WALS*, *D-Place*, *Ethnologue* and *Glottolog*. Our preliminary results find partial support for the above correlations. In general, albeit with some exceptions, Type A languages tend towards more complex morphosyntax and greater expressive power in certain domains, although also towards more complex phonological inventories, while Type B languages tend towards more complex morphology.

Next, we hypothesize that this crosslinguistic variation entails differential involvement of declarative versus procedural memories. Procedural memory subserves the acquisition of compositional, automated, rule-governed (grammatical) aspects of language, while declarative memory typically subserves vocabulary learning and irregular phenomena across domains, including memorized, opaque, formulaic language (e.g. idioms and proverbs) (Ullman, 2004; 2015; Heyselaar et al., 2017; Elyoseph et al., 2020, for impairments). While both memory systems are essential for language (with partly overlapping/redundant functions), and while both language types certainly rely on both memories, our hypothesis is that predominantly Type A languages rely more on procedural memory, while predominantly Type B languages, in comparison, rely more on declarative memory. For testing this, we are conducting standard cognitive experiments measuring the relative strengths of the two memory types with speakers of Type A vs. Type B languages. Also, because these two types of memories depend on brain regions whose emergence is genetically guided during development, another way of testing our hypothesis is by seeking correlations between the Type A/Type B linguistic distinction, and the frequency in the population of the candidate gene alleles supporting different memory types. Various genes have been found to play a role in declarative memory, e.g., *BDNF* and *APOE* (Ullman, 2015; Henke, 2010; Squire & Wixted, 2011; Eichenbaum, 2012), as opposed to procedural memory, e.g. *FOXP2*, *PPP1R1B* and *DRD2* (Packard, 2008; Doyon et al., 2009; Ashby et al., 2010; Eichenbaum, 2012). Since cognitive biases can be linked to (epi)genetic modifications, any differential reliance with respect to the two types of memories is expected to be detectable in differences in the allele frequencies of specific genes. At present, we have found differences between speakers of Type A and Type B languages with regards to genes involved in synapse organization subserving relevant brain functions.

This approach provides a tangible way to engage the neurobiological and genetic underpinnings of language variation, identifying specific and testable implications for the role of both cultural and genetic factors in language evolution, where they are seen as engaged in a feedback loop, with each reinforcing the other. As a bonus, our proposal sheds novel light on the long-standing linguist’s puzzle (and controversy), where researchers often report trade-offs in complexity among different linguistic domains (e.g. Sampson, Gil & Trudgill, 2009).

## References

Ashby, F. G., Turner, B. O., & Horvitz, J. C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. *Trends in Cognitive Sciences, 14* (5), 208–215.

Doyon, J., Bellec, P., Amsel, R., Penhune, V. B., Monchi, O., Carrier, J., Lehéricy, S., Benali, H. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behavioural Brain Research, 199,* 61–75.

Eichenbaum, H. (2012). *The Cognitive Neuroscience of Memory: An Introduction* (2nd ed.). Oxford: Oxford University Press.

Elyoseph, Z., Mintz, M., Vakil, E., Zaltzman, R., Gordon, C.R. (2020). Selective procedural memory impairment but preserved declarative memory in Spinocerebellar Ataxia Type 3. *The Cerebellum* 19:226–234. https://doi.org/10.1007/s12311-019-01101-w

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience, 11,* 523–532.

Heyselaar, E., Segaert, K., Walvoort, S. J.W., Kessels, R. P.C., Hagoort, P. (2017). *The role of nondeclarative memory in the skill for language: Evidence from syntactic priming in patients with amnesia. Neuropsychologia. 101,* 97-105. doi:10.1016/j.neuropsychologia.2017.04.033

Packard, M. G. (2008). Neurobiology of procedural learning in animals. In J. H. Byrne (Ed.), *Concise learning and memory: The editor's selection* (pp. 341–356). London: Elsevier Science and Technology.

Sampson, G., Gil, D., and Trudgill, P. (Eds.) (2009). *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press.

Squire, L. R., and Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience, 34,* 259–288.

Ullman, M. T. (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition 92 (1–2),* 231–270.

Ullman, M. T. (2015). The declarative/procedural model: a neurobiologically motivated theory of first and second language. In Van Patten, B., and Williams, J. (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd ed.) (pp. 135-158). London and New York: Routledge.

Wray, A., Grace, G.W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua 117*, 543–578.

# CONTINUITY OF EMOTION AND AGGRESSION IN LANGUAGE EVOLUTION

ANTONIO BENÍTEZ-BURRACO[*1], LJILJANA PROGOVAC[2]

[*]Corresponding Author: abenitez8@us.es
[1]Department of Spanish, Linguistics & Theory of Literature, University of Seville, Spain
[2] Department of English (Linguistics), Wayne State University, Detroit, USA

While propositional language has been at the center of language evolution theorizing, it faces a problem of discontinuity when it comes to other primates' communication. Our main claim is that continuity should not be abandoned, as it reveals itself in a gradualist approach that ties together the evolution of language/syntax with the management of emotion/aggression, where the two are engaged in a co-evolutionary feedback loop. We rely on the finding that the same mechanism of dense neuronal connectivity between cortical and subcortical structures, implicated in the suppression of reactive aggression, is also responsible for cross-modality (metaphoricity) and for syntactic processing (Benítez-Burraco and Progovac 2021).

In this respect, early forms of language, with simplest grammars and highly concrete, imageable vocabulary have been proposed to be more visceral (and linked to older processing strategies) than typical constructions of modern languages, in which even imageable, emotive words get buried under the layers of abstract syntactic categories (Progovac et al. 2018). Even today, metaphorical, imageable vocabulary has a higher emotional impact than literal language, relying more on the limbic system (e.g. Bohrn et al. 2012). These earliest stages of language would have shown more continuity with animal communication (i.e. vocalizations), which are highly emotional (Darwin 1872; Code 2011).

We tie our proposal to the history of genetic and biological changes associated with changes in the management of aggression, which resulted in physical, behavioral and cognitive changes in our species, and which are sometimes referred to as human self-domestication (HSD) (Hare et al., 2012). The period before 200 kya saw high levels of reactive aggression and hominins at that stage probably relied on brief (even single word) emotion-driven outbursts, such as commands, warnings, and threats (*Watch out! Run! Move*!) The fossil record suggests that features of HSD, and implicitly of reactive aggression, were gradually decreasing between 200-100 kya (Leach, 2003; Zollikofer and Ponce de León, 2010), which arguably coincided with the emergence of simple (two-slot) grammars, capable of merging only one verb and one noun (e.g. *scatter-brain*; *hunch-back*; *turn-skin*; *cry-baby*; Progovac, 2015). While they would have been truly beneficial in general, these highly metaphorical compositions also turn out to be especially well-suited for insult/verbal aggression, and as such they would have been highly adaptive at this stage in evolution, significantly contributing to replacing physical aggression/contest with verbal aggression/cognitive contest. This further reinforced the trend in selection towards increased control of subcortical networks by cortical devices, contributing further to suppression/inhibition of reactive aggression, and ultimately, to HSD features. Based on the fossil record, HSD reached its peak roughly between 100-50 kya and 10kya (Cieri et al., 2014), at the time when fully-fledged grammars emerged and referential uses of language started to become more important.

Neurobiological and neuropsychiatric findings provide support for our proposal. Reactive (physical) aggression typically implicates the limbic system (Rolls 2015 for review), the striatal regions, and parts of the cortex (Dolan et al. 2002; Yang et al. 2009; Boccardi et al. 2011). The processing of aggressive language, such as swear words and profanity, shows a clear overlap/continuity with physical aggression, implicating the basal ganglia, limbic structures, thalamus, and the right hemisphere (e.g. Code 2011). Disorders which sometimes result in uncontrolled swearing/profanity (e.g. Tourette's Syndrome) also involve enhanced physical aggression, attributable to a basal-limbic connection dysfunction (e.g. Ganos et al. 2013).

Overall, our proposal supports the view that human evolution saw a gradual shift from raw, emotional language (with rudimentary syntax), to structurally highly complex forms of language, which can be quite detached from emotion, and which arguably correlated with the relatively recent rise in proactive/premeditated aggression (Wrangham 2018). Less visceral, more rational uses of language of course open new possibilities for dishonest signaling. The gradual progression in human evolution from reactive aggression (driven by raw emotion), to proactive/premeditated aggression (driven by more detached reasoning), is itself a transition from more emotional to more rational behavior.

## Selected References

Benítez-Burraco A. and Progovac L. (2021). Language evolution: examining the link between cross-modality and aggression through the lens of disorders. *Philos Trans R Soc Lond B Biol Sci*. 376(1824):20200188

Boccardi, M., Frisoni, G. B., Hare, R. D., Cavedo, E., Najt, P., Pievani, M., et al. (2011). Cortex and amygdala morphology in psychopathy. *Psychiatry Res.* 193, 85–92. doi: 10.1016/j.pscychresns.2010.12.013

Bohrn, I., Altmann, U. and Jacobs, A. M. (2012). Looking at the brains behind figurative language - A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia* 50, 2669-2683. doi: 10.1016/j.neuropsychologia.2012.07.021

Cieri, R. L., Churchill, S. E., Franciscus, R. G., Tan, J., and Hare, B. (2014). Craniofacial feminization, social tolerance, and the origins of behavioral modernity. *Curr. Anthropol*. 55:419–43.

Code, C. (2011). Nonfluent aphasia and the evolution of proto-language. *J. Neurolinguistics* 24, 136–144. doi: 10.1016/j.jneuroling.2009.12.007

Darwin, C.R. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray. 1st edition.

Dolan, M., Deakin, W. J., Roberts, N., and Anderson, I. (2002). Serotonergic and cognitive impairment in impulsive aggressive personality disordered offenders: are there implications for treatment? *Psychol. Med.* 32, 105–117. doi: 10.1017/S0033291701004688

Ganos C, Münchau A, Bhatia K. (2014). The semiology of tics, Tourette's, and their associations. *Mov. Disord. Clin. Pract*. 1, 145–153. doi:10.1002/mdc3.12043

Hare, B., Wobber. V., Wrangham, R. (2012). The self-domestication hypothesis: Evolution of bonobo psychology is due to selection against aggression. *Animal Behavior* 83.3: 573-585.

Leach, H.M. (2003). Human domestication reconsidered. *Curr. Anthropol.*, 44, 349–368.

Progovac L. (2015). *Evolutionary Syntax*. Oxford, UK: Oxford University Press.

Progovac, L., Rakhlin, N., Angell, W., Liddane, R., Tang, L., and Ofen, N. (2018). Neural correlates of syntax and proto-syntax: an fMRI study. *Front. Psychol.* 9:2415. doi: 10.3389/fpsyg.2018.02415

Rolls, E. T. (2015). Limbic systems for emotion and for memory, but no single limbic system. *Cortex* 62, 119–157. doi: 10.1016/j.cortex.2013.12.005

Yang, Y., Raine, A., Colletti, P., Toga, A. W., and Narr, K. L. (2009). Abnormal temporal and prefrontal cortical gray matter thinning in psychopaths. *Mol. Psychiatry* 14, 561–562. doi: 10.1038/mp.2009.12

Wrangham R. W. (2018). Two types of aggression in human evolution. *PNAS*, 115(2), 245–253. https://doi.org/10.1073/pnas.1713611115

Zollikofer, C. P. and Ponce de León, M. S. (2010). The evolution of hominin ontogenies. *Semin. Cell Dev. Biol.* 21(4), 441–452.

# COMPLEX SOUND SEQUENCES IN ITERATED LEARNING: CLUSTERS CAN BECOME MORE STABLE WHEN THEY SIGNAL MORPHOLOGICAL BOUNDARIES

IRENE BÖHM*[1] and NIKOLAUS RITT[1]

*Corresponding Author: irene.boehm@univie.ac.at
[1]Department of English & American Studies, University of Vienna, Vienna, Austria

We report a spoken artificial iterated language learning experiment in which we investigated the transmission of final consonant clusters. Our main research question was how well clusters that are exclusively produced through suffixation (such as the /gz/ in *egg+s*) are transmitted in comparison to (a) exclusively stem-internal clusters (such as the /lk/ in *hulk*), and (b) clusters that occur both within stems and across boundaries (such as /ks/ in *fix* and *kick+s*).

In comparison to consonant-vowel sequences, consonant clusters count as highly 'marked' or 'dispreferred' in many linguistic theories (cf. Clements & Keyser, 1983; Prince & Smolensky, 2004; Dziubalska-Kołaczyk, 2002, 2019). They are difficult both to perceive and to produce, cross-linguistically rare (Maddieson, 1999, 2003), and historically unstable (e.g. Lutz, 1988). This implies that they are selected against in cultural evolution (Ritt, 2004; Christiansen & Chater, 2008). However, consonant clusters appear to be more frequent and more stable when they are 'morphotactic', i.e. when they span morpheme boundaries, as in English /gz/ in *egg+s*. It has been hypothesized that this is because they indicate morphological complexity and facilitate the processing of complex word forms (Dressler & Dziubalska-Kołaczyk, 2006).

From this reasoning an interesting question arises about final clusters that occur both across morpheme boundaries and within word stems, such as English /nd/ in simple *find* and complex *dine+d*, for example. In contrast to unambiguously morphotactic clusters like the /gz/ in *egg+s*, they do not reliably signal complexity and may even impede processing (Post et al., 2008). This suggests that they should be difficult to transmit and be unstable in cultural language evolution. On the other hand, however, it has also been argued that frequent morphotactic clusters may support the emergence and the stability of stem-internal homophones through analogy (Hogg & McCully, 1987). Evidence for both hypotheses has been found in processing studies (e.g. Celata et al., 2015) and diachronic research (e.g. Baumann et al., 2016, 2019). We investigate transmission of morphotactic, stem-internal, and ambiguous clusters by means of a spoken iterated artificial language learning experiment.

Following the basic design of such experiments (Kirby, Cornish & Smith, 2008), we asked participants to learn words from an artificial miniature language and to reproduce them from memory. The outputs of ‘first-generation’ participants served as learning input for a second ‘generation’. This procedure was then repeated over ten participant-generations. In our specific experiment, thirty Austrian German native speakers learned a set of singular and plural nouns, some of which contained a highly marked plosive+/k/ cluster (e.g. /zʊtk/). They were trained on eleven nouns, each of which comprised an abstract image and a spoken label, and subsequently asked to reproduce the sound labels corresponding to each image. To address our research question, we set up three conditions. In condition one, all clusters occurred within singular stems (e.g. SG /zʊtk/ – PL /zʊtk+ɒ/). In condition two, all clusters were produced by adding the plural suffix {-/k/} to stems (e.g. SG /zʊt/ – PL /zʊt+k/). Finally, in condition three, the clusters occurred both within singular stems and across stem-suffix boundaries.

In our analysis, we compared the transmission of lexical, morphotactic, and ambiguous clusters. Our results showed that /tk/ clusters remained stable only when they signaled plural suffixation unambiguously, i.e. in condition (ii), where they were successfully transmitted across all ten participant generations. In contrast, both exclusively stem-internal clusters (condition i) and ambiguous clusters (condition iii) were reduced and/or lost early in transmission. Quantitatively, the significant interaction between cluster transmission and morphotactic condition was substantiated in multivariate regression models and a Pearson's correlation test.

We take our data as support for the position that morphotactic cross-boundary clusters are selected in cultural evolution, but only when they signal complexity unambiguously. If they have stem-internal homophones, on the other hand, this does not seem to be the case. Instead, ambiguous clusters seem to be as unstable, as clusters that occur only within stems and have no morphological signaling function at all. However, the possibility that cluster homophones may also support each other cannot be ruled out, since the relevant processes may occur in early stages of first language acquisition, while our experiment involved only adult participants.

## References

Baumann, A., Prömer, C., & Ritt, N. (2019). Word form shapes are selected to be morphotactically indicative. *Folia Linguistica Historica, 40*(1), 129–151.

Baumann, A., Ritt, N., & Prömer, C. (2016). Diachronic dynamics of Middle English phonotactics provide evidence for analogy effects among lexical and morphonotactic consonant clusters. *Papers in Historical Phonology, 1,* 50–75.

Celata, C., Korecky-Kröll, K., Ricci, I., & Dressler, W. U. (2015). Phonotactic processing and morpheme boundaries: Word-final/Cst/clusters in German. *Italian Journal of Linguistics, 27,* 185-110.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*(5), 489-509.

Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Dressler, W. U., & Dziubalska-Kołaczyk, K. (2006). Proposing morphonotactics. *Wiener Linguistische Gazette, 73,* 69–87.

Dziubalska-Kołaczyk, K. (2002). *Beats-and-Binding Phonology.* Frankfurt am Main: Peter Lang.

Dziubalska-Kołaczyk, K. (2019). On the structure, survival and change of consonant clusters. *Folia Linguistica, 40*(1), 107–127.

Hogg, R. M., & McCully, C. B. (1987). *Metrical phonology: A course book.* Cambridge: Cambridge University Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681–10686.

Lutz, A. (1988). On the historical phonotactics of English. In D. Kastovsky and G. Bauer (Eds.), *Luick revisited: Papers read at the Luick-Symposium at Schloß Liechtenstein* (pp. 221-239). Tübigen: Narr.

Maddieson, I. (1999). In search of universals. *ICPhS99, 3,* 2521-2528.

Maddison, I. (2013). Syllable structure. In M. S. Dryer and M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/12.

Post, B., Marslen-Wilson, W. D., Randall, B., & Tyler, L. K. (2008). The processing of English regular inflections: Phonological cues to morphological structure. *Cognition, 109,* 1–17.

Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in Generative Grammar.* Malden, MA: Wiley Blackwell.

Ritt, N. (2004). *Selfish sounds and linguistic evolution: A Darwinian approach to language change.* Cambridge: Cambridge University Press.

# EFFECTS OF MEANING NEGOTIATION ON CORRECTNESS, GESTURE COUNT AND TIMESPAN IN A GAME OF CHARADES

Monika Boruta*, Marek Placiński*, Marta Sibierska*, and Przemysław Żywiczyński*

Corresponding Author: monika.feba@vp.pl
*Nicolaus Copernicus University in Torun, Poland

Meaning negotiation, most commonly described in language learning and acquisition research (Long 1981; Pica, Young & Doughty 1987), facilitates mutual understanding of interlocutors and influences efficiency of communication (Rees 1998). Negotiation of meaning is based on repetition of words, usage of synonyms and restructuring phrases in a communicative situation in order to enable the parties to work out common means of expression (Rees 1998; Dobao & Martinez 2007, SL: Zeshan 2015). If speech fails, one can resort to the use of gesture and pantomime (Dobao & Martinez 2007, SL: Zeshan 2015) – which are often said to be a communicative universal, that can be produced „on the fly" (Arbib 2018; Zlatev et al. 2020, Silva et al. 2020). We analysed the process of meaning negotiation with respect to three factors: (i) communicative success, (ii) gesture count/pantomime, and (iii) timespan, in a mute game of charades. Following insights from gesture and sign language research, we assumed that meaning negotiation causes not only greater understanding over time, but – what we focus our analysis on – a reduction of (pantomimic) gestures or signs and their conventionalisation and the time of message presentation (Earley 1999; Dachkovsky et al. 2018, Mineiro et al. 2021; Namboodiripad et al. 2016). To test this assumption, we used a dataset of video recordings of pantomime, collected in a study based on the referential task paradigm known from Experimental Semiotics (see e.g. Fay et al. 2010).

The recordings include interactions in 26 pairs of participants, who played a mute game of charades consisting of 4 rounds and based on written input. We analysed the following factors in the interactions: (I) communicative success, operationalised as the number of correct responses in each round; (II) meaning unit count, operationalised as the number of lexical items from the input re-enacted by means of gesture and pantomime in each round; (III) timespan of each round, given in seconds. We focused on the differences in the three factors in round 1, before the negotiation, and in round 4, after the negotiation.

In order to verify hypotheses, we performed three separate regression analyses for each outcome variable (Correctness – Fig. 1, Gesture Count – Fig. 2, Time – Fig. 2). Gesture count and Time were standardised before being entered into their respective models in R. In line with our assumptions, the participants negotiated the meaning throughout the game and were making significantly more mistakes in round 1 than they did in round 4; the time they spent on presenting their re-enactments in round 1 compared to round 4 was also significantly shorter; however, the number of gestures remained at the same level.

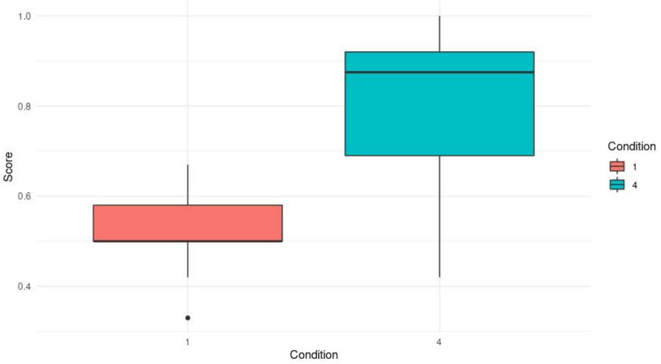


Figure 1. Box plots for correctness for rounds 1 v. 4.

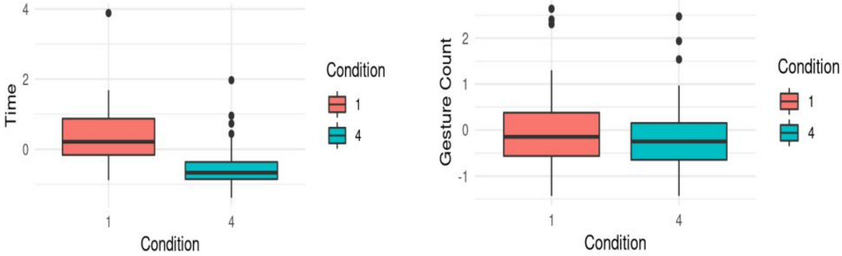


Figure 2. Box plots: time span for rounds 1 v. 4 and gesture count for rounds 1 v. 4.

In the discussion, we propose explanations for such a result - we refer to research in sign/language change over time, language evolution, communicative pressures, and meaning negotiation processes. We also discuss the results in light of a more detailed description of the participants’ gestural choices and strategies between the rounds.

## Acknowledgements

This research was funded by the Polish National Science Centre (NCN) under the agreement UMO-2017/27/B/HS2/00642.

## References

Arbib, M. A. (2018). In support of the role of pantomime in language evolution. Journal of Language Evolution, 3(1), 41-44.

Dachkovsky, S., Stamp, R., & Sandler, W. (2018). Constructing complexity in a young sign language. Frontiers in Psychology, 9, 2202.

Dobao, A. M. F., & Martínez, I. M. P. (2007). Negotiating meaning in interaction between English and Spanish speakers via communicative strategies. Atlantis, 87-105.

Earley, K. (1999). Historical Changes in Sign Language-a Comparison of Spoken and Sign Language Change.[Thesis, online access 14.02.2022].

Fay, N. et al. (2010). The interactive evolution of human communication systems. *Cognitive Science*, 34(3), 351–386.

Long, M. H. 1981. “Input, interaction and second language acquisition”. In Winitz, H. (ed.), Annals of the New York Academy of Sciences: Native Language and Foreign Language Acquisition 379, 259-278

Mineiro, A., Báez-Montero, I. C., Moita, M., Galhano-Rodrigues, I., & Castro-Caldas, A. (2021). Disentangling pantomime from early sign in a new sign language: window into language evolution research. Frontiers in Psychology, 12, 1130.

Namboodiripad, S., Lenzen, D., Lepic, R., & Verhoef, T. (2016). Measuring conventionalization in the manual modality. Journal of Language Evolution, 1(2), 109-118.

Rees, D. (1998). The negotiation of meaning in EFL learning in the language laboratory. ASp. la revue du GERAS, (19-22), 283-309.

Pica, T., R. Young & C. Doughty. (1987). “The impact of interaction on comprehension”. TESOL Quarterly 21/4, 737-758.

Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. Royal Society open science, 7(1), 182056.

Zeshan, U. (2015). “Making meaning”: Communication between sign language users without a shared language. Cognitive Linguistics, 26(2), 211-260.

Zlatev, J., Żywiczyński, P., Wacewicz, S. (2020). Pantomime as the original human-specific communicative system. *Journal of Language Evolution*, 5(2), 156–174, doi: 10.1093/jole/lzaa006.

# SEQUENTIAL DYNAMICS IN THE VOCAL SYSTEM OF COMMON MARMOSETS (*CALLITHRIX JACCHUS*)

Alexandra B. Bosshard[1,2,*], Balthasar Bickel[1,2], Paola Merlo[3], Judith Burkart[2,4] and Simon W. Townsend[1,2,5]

[*]Corresponding Author: alexandra.bosshard@uzh.ch
[1]Department of Comparative Language Science, University of Zurich, Zurich,
[2]Center for the Interdisciplinary Study of Language Evolution (ISLE),
University of Zurich, Zurich, Switzerland
[3]Department of Linguistics University of Geneva, Geneva, Switzerland
[4]Department of Anthropology, University of Zurich, Zurich, Switzerland
[5]Department of Psychology, University of Warwick, Coventry, UK

Over the last 20 years there has been growing interest in the combinatorial abilities of animals, namely the propensity to sequence context-specific calls (see e.g., Arnold & Zuberbühler, 2006; Berthet et al., 2019; Coye, Ouattara, Zuberbühler, & Lemasson, 2015; Coye, Zuberbühler, & Lemasson, 2016), given the light such data can shed on the evolutionary progression of our own combinatorial communication system (Leroux & Townsend, 2020; Townsend, Engesser, Stoll, Zuberbühler, & Bickel, 2018).

However, most studies to date have concentrated on the concatenation of two call units (bigrams) (see e.g., Arnold & Zuberbühler, 2006; Coye, Ouattara, Zuberbühler, & Lemasson, 2015; Coye, Zuberbühler, & Lemasson, 2016; Collier et al., 2020) with such structures even being argued to represent the limit of animal combinatorial abilities (see e.g., Miyagawa & Clarke, 2019). Whether animal combinatorics lie beyond the two call level remains less explored, yet is critical for understanding how similar human and animal vocal systems are in their sequential dynamics.

We revisit this question through investigating the formation and internal structuring of call sequences in the highly social and cooperatively breeding common marmoset (*Callithrix Jacchus*). We recorded the vocal output of eight individuals (four pairs housed at the University of Zurich primate station, Switzerland, resulting in four female and four male subjects) via focal sampling.

This led to a corpus of 8600 calls, equalling 5800 call sequences (from unigram to 9-gram sequences), generated by the eight different call types present in the common marmoset vocal system. Through borrowing methods traditionally applied in computational linguistics, including simple frequency values, finite state automata (FSA), superordinate Markov models (MMs) (for an application in song systems see Honda and Okanoya, 1999; Hosino and Okanoya, 2000; Kershenbaum et al., 2014; Sainburg et al., 2019; Suzuki, Buck and Tyack, 2006; Kershenbaum et al., 2014) and collocation analysis, we show that common marmosets reliably concatenate vocalisations into larger structures of between two and nine calls (trigrams being the most common length of combinations).

Furthermore, using FSA and MMs, we identified sequential relationships between calls in combinations that point towards potential internal structuring. For example, trigram MMs highlighted a non-adjacent-like edge structure (a dependency between the first and last call element of a sequence (see Endress et al., 2009; Endress et al., 2010 and Grama, Wijnen & Kerkhoff, 2013)), whereby a trigrammic sequence starting with a food call always terminates with another food call in the last position. Multiple distinctive collocation analysis confirmed the food-call edge structure for three-call combinations and also four-call combinations. In addition, joint probability showed that the non-adjacent-like food-edge structure was present in 86% of all trigrams and 66% of all 4-grams.

While previous comparative research has identified non-adjacent structuring (structures that cannot be captured by more simple Markovian dynamics) in animal song systems (e.g., Sainburg et al. 2019; Suzuki, Buck and Tyack, 2006; Kershenbaum et al., 2014), this has not yet been shown in a non-human animal context-specific call system as is the case for the common marmosets' vocal communication system.

MMs and collocation analyses additionally provided some evidence for simple nestedness in common marmosets' call structures. For example, the bigram Tsk-Ek, which was identified by joint probability values and by the collocation analyses as one of the most common two-call units, also shows up in trigrammic combinations and when it does, it is reliably followed by a contact call giving rise to a "bigram-in-bigram" structuring.

We argue that call combinations in animals may be more sophisticated than previously thought and that the current dearth of evidence for larger call combinations in non-human animals could well be due to a lack of application of objective measurements to capture the sequential dynamics of structures at the call combination level.

**References**

Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*, *441*(7091), 303–303.

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. Science Advances, 5(5), eaav3991.

Collier, K., Radford, A. N., Stoll, S., Watson, S. K., Manser, M. B., Bickel, B., & Townsend, S. W. (2020). Dwarf mongoose alarm calls: investigating a complex non-human animal call. Proceedings of the Royal Society B, 287(1935), 20192514.

Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. Proceedings of the Royal Society B: Biological Sciences, 282(1807), 20150265.

Coye, C., Zuberbühler, K., & Lemasson, A. (2016). Morphologically structured vocalizations in female Diana monkeys. Animal Behaviour, 115, 97–105.

Endress A.D., Carden S., Versace E. & Hauser M.D. (2010). The apes' edge: positional learning in chimpanzees and humans. Animal cognition, 13(3), 483-495.

Grama I., Wijnen F. & Kerkhoff A. (2013). Constraints on non-adjacent dependency-learning: Distance matters an artificial grammar learning study with adults. In Boston University Conference on Language Development 37 online proceedings supplement, Boston, MA.

Honda, E., & Okanoya, K. (1999). Acoustical and syntactical comparisons between songs of the white-backed munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica). Zoological Science, 16(2), 319-326.

Hosino, T., & Okanoya, K. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. Neuroreport, 11(10), 2091-2095.

Hurford, J. R. (2012). The origins of grammar: Language in the light of evolution II (Vol. 2). Oxford University Press.

Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: Not the Markov chains we thought they were. Proceedings of the Royal Society B: Biological Sciences, 281(1792), 20141370.

Leroux, M., & Townsend, S. W. (2020). Call combinations in great apes and the evolution of syntax. Animal Behavior and Cognition, 7(2), 131–139.

Miyagawa, S., & Clarke, E. (2019). Systems underlying human and old world monkey communication: One, two, or infinite. *Frontiers in psychology*, 1911.

Sainburg, T., Theilmann, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. Nature Communications, 10(1), 1-11

Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. The Journal of the Acoustical Society of America, 119(3), 1849–1866.

Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. PLoS Biology, 16(8), e2006425.

# DOES A COMMUNICATIVE CONTEXT PRODUCE A PREFERENCE FOR EARLY NEGATION? AN ARTIFICIAL LANGUAGE LEARNING STUDY

Danielle Burgess*[1]

*Corresponding Author: dlburge@umich.edu
[1]Linguistics Department, University of Michigan, Ann Arbor, USA

Tendencies for negation to appear early in the sentence have been observed by scholars working in typology (Dryer, 1988), creole language emergence (Schneider, 2000), and language acquisition (Kim & Yun, 2013; Dimroth, 2010). Some scholars have proposed that a tendency for early negation marking may be motivated by considerations of audience design in communication, because delaying the negator could increase misunderstanding or comprehension difficulty when the sentence up to but not including the negative is the opposite of the intended meaning (Dryer, 1988: 102; Horn, 1989: 449). However, such proposals have not yet been widely investigated. Furthermore, evidence for an early negation preference in L1 and L2 acquisition is complicated by confounds of difficulty, complexity, or frequency of negative constructions and co-occurring elements such as, for example, verb-raising (Dimroth 2010).

The present study utilizes an artificial language learning paradigm to verify the presence of a preference for earlier negation in a language that allows for free variation between preverbal and postverbal negation. Participants were also paired to take part in a director-matcher paradigm to investigate whether a preference for early negation could be plausibly driven by the specific pressures of communication, as opposed to learning and production in isolation.

54 English-speaking participants were exposed to a miniature language that contained a mixture of preverbal and postverbal negation (NegVSO and VNegSO). Participants were divided equally among three experimental conditions corresponding to different proportions of preverbal negation ordering they were exposed to during training (25%, 50%, and 75%). After training, participants first produced sentences in a non-communicative context by

describing images withheld from the training set. Then, participants were paired with another learner (trained on the same input proportions) to perform a director-matcher task, followed by a final solo retest.

Across conditions and production phases, participants produced more preverbal negation on average compared to the proportion of preverbal negation in the input (see Figure 1). Statistical analysis using mixed effects logistic regression revealed that the use of the majority order in training was significantly lower in the Majority VNeg condition than the Majority NegV condition ($\beta = -0.97$, $SE = 0.35$, $p = 0.006$). There was also a significant interaction indicating an additional effect on the difference in use of the majority order between the Majority NegV and Majority VNeg conditions in the interaction stage, compared to pre-interaction production ($\beta = -0.36$, $SE = 0.09$, $p < 0.001$). Subsequent pairwise comparisons revealed that within each condition, the only significant differences in use of NegV order between production stages were within the Majority NegV training condition, where there was significantly more NegV use in the interaction stage ($\beta = -1.22$, $SE = 0.29$ $p = 0.001$) and post-interaction stage ($\beta = -1.10$, $SE = 0.28$ $p = 0.003$) compared to pre-interaction production.

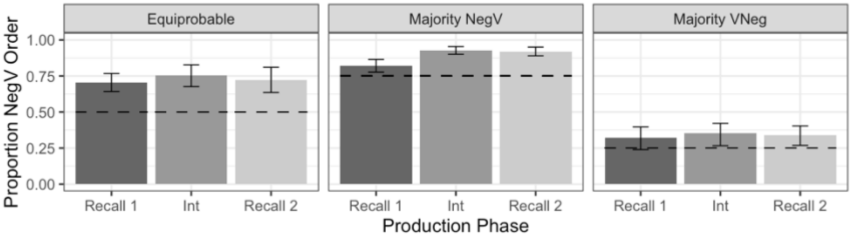


Figure 1. Plot showing participants' average proportion of NegV production. The dashed lines indicate the proportion of NegV order seen in training. Error bars represent standard error.

These results verify the existence of a production bias consistent with preverbal negation preferences in adult English speakers learning and producing a language which contains both preverbal and postverbal negation, consistent with preferences observed in, for example, naturalistic L2 learning of Korean (Kim & Yun, 2013). Further research is necessary to test whether this preference extends to speakers of postverbal negation languages like Japanese, though note that in studies of naturalistic L2 acquisition, even learners with L1s that use postverbal negation have also shown some preference for preverbal negation patterns (Kim & Yun, 2013; Hyltenstam 1977). Looking at the role of interaction, the director-matcher task failed to reliably boost NegV use compared to production in isolation prior to interaction, which does not support the idea that a communicative context induces a greater preference for preverbal negation in general.

## Acknowledgements

This research was supported by the NSF Graduate Research Fellowship (grant number DGE-1841052) and a Rackham Graduate School research grant.

## References

Dimroth, C. (2010). The acquisition of negation. In L. R. Horn (Ed.), *The Expression of Negation*. Berlin, New York: De Gruyter Mouton.

Dryer, M. S. (1988). Universals of negative position. M. Hammond, E. Moravcsik, & J. Wirth (Eds.), *Studies in Syntactic Typology*. Amsterdam: John Benjamins.

Horn, Laurence R. (1989). *A natural history of negation*. Chicago: University of Chicago Press.

Hyltenstam, K. (1977). Implicational patterns in interlanguage syntax variation. *Language Learning* 27, 383–410.

Kim, H. K. & Yun H. (2013). Examining the Influence of L1 on Learning Korean Negation Forms. *Foreign Languages Education*, 20. 415–434.

Schneider, E. W. (2000). Feature diffusion vs. contact effects in the evolution of new Englishes: A typological case study of negation patterns. *English World-Wide* 21, 201–230.

# PROSOCIALITY IN SWARM ROBOTICS: A MODEL TO STUDY SELF-DOMESTICATION AND LANGUAGE EVOLUTION

Nicolas Cambier*[1], Roman Miletitch[2], Antonio Benítez Burraco[3], and Limor Raviv[4,5,6]

*Corresponding Author: n.p.a.cambier@vu.nl
[1]Computational Intelligence Group, Vrije Universiteit Amsterdam, The Netherlands
[2]IRIDIA, Université Libre de Bruxelles, Belgium
[3]University of Seville, Spain
[4]LEADS group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[5]Centre for Social, Cognitive and Affective Neuroscience, University of Glasgow, UK
[6]Artificial Intelligence Lab, Department of Computer Science, Vrije Universiteit, Brussels, Belgium

## 1. Introduction

Swarm Robotics, which studies collective behaviors of large populations of interacting robots with simple embodied cognition, is an ideal testbed for studying the cultural evolution of language. Beyond simple communication processes, past experiments in swarm robotics have explored the potential of language games in an embodied agent context (Trianni, De Simone, Reina, & Baronchelli, 2016; Cambier, Frémont, & Ferrante, 2017; Cambier, Albani, Frémont, Trianni, & Ferrante, 2021; Miletitch, Reina, Dorigo, & Trianni, 2019).

The current study expands on existing simulations in order to explore a dominant theory in evolutionary linguistics, namely, that the evolution of present-day languages might have resulted in part from the self-domestication of the human species (Thomas & Kirby, 2018; Benítez-Burraco & Progovac, 2020). According to this human self domestication (HSD) hypothesis, humans' distinctiveness is, to a large extent, the outcome of an evolutionary process similar to animal domestication (Hare, 2017). At the heart of this view lies the idea that HSD resulted in less aggressive individuals, who are more prone to interact with others (and particularly, with their kin, but also with strangers). This increased prosociality and reduced aggressiveness would have in turn promoted more social contacts within a community, and supported the emergence of more sophisticated forms of language (Tamariz & Kirby, 2016; Steels, 2017). Our goal is to create an embodied model of the effects of prosociality on the formation of language using swarm robots, in order to investigate the self-domestication hypothesis and the process of language evolution in general.

Current swarm robotics models, however, lack several crucial features that are considered prerequisites for this process. Swarm robots are typically collaborative, but homogeneous, and have little to no memory, not to mention social memory (who did what to whom). In order to mimic the effects of self-domestication properly, an evolutionary advantage for prosociality first needs to be introduced, and robots need to be treated as distinct individuals. To this end, we designed a novel version of the naming game with swarm robots.

## 2. The current model

In our model, robots in multiple nests are engaged in a foraging task (i.e., gathering resources in their environment) while playing a naming game (Steels & Loetzsch, 2012). Foraging and resource collection are typical animal activities, which has been widely studied in the field of swarm robotics (e.g. (Miletitch, Dorigo, & Trianni, 2018)) and constitutes an obvious marker of evolutionary fitness, which is highly relevant to an investigation of the HSD. Crucially, we include two novel features in the model: (1) robot individuation: robots have a partner-specific memory, keeping track of the outcomes of past interactions with specific robots; (2) parametrizable prosociality: robots' tendency to interact is based on (a) an innate factor and (b) experience: successful communication between robots reduces their aggression toward each other and increases their chances of interacting again.

We examine the evolution of communicative alignment and foraging behavior within and between nests as a function of prosociality and geographic distance, which are two key factors accounting for language diversity and impacting on language complexity (Padilla-Iglesias, Gjesfjeld, & Vinicius, 2020; Bickel & Nichols, 2009). First, we ask whether increased prosociality leads to more efficient foraging, and whether this pattern is affected by the geographical proximity of nests (seeing as closer nests result in more competition over the same resources). Second, we ask whether our manipulation of initial prosociality affects classic convergence patterns (i.e., where all robots end up aligning on the same word variant), leading to differential divergence depending on nest and past history. That is, does initial prosociality affect the degree and/or speed of convergence within and between nests, and does it result in a distinction between in-group robots (belonging to the same nest) and out-group robots (belonging to different nests)?

## 3. Results

First, we show that these manipulations lead to the formation of a classic "in-group bias" where robots favor interaction with some robots over others - a bias which is highly common in social animals in nature but that was so far absent from swarm robotics models. Second, we observe that higher prosociality values result in the collection of more resources, potentially indicating an evolutionary advantage. Finally, we show that prosociality modulates the effect of physical distance on lexical convergence, such that low values of innate prosocially lead to more stable sub-swarm divergence, even in relative proximity (i.e., different nests robustly converge on different lexical variants despite being close to each other).

## References

Benítez-Burraco, A., & Progovac, L. (2020). A four-stage model for language evolution under the effects of human self-domestication. *Language & Communication*, *73*, 1–17.

Bickel, B., & Nichols, J. (2009). The geography of case. *The Oxford handbook of case*, *479*, 493.

Cambier, N., Albani, D., Frémont, V., Trianni, V., & Ferrante, E. (2021). Cultural evolution of probabilistic aggregation in synthetic swarms. *Applied Soft Computing*, *113*, 108010.

Cambier, N., Frémont, V., & Ferrante, E. (2017). Group-size regulation in self-organised aggregation through the naming game. In *International symposium on swarm behavior and bio-inspired robotics (swarm 2017).* Kyoto, Japan.

Hare, B. (2017). Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual review of psychology*, *68*, 155–186.

Miletitch, R., Dorigo, M., & Trianni, V. (2018). Balancing exploitation of renewable resources by a robot swarm. *Swarm Intelligence*, *12*(4), 307–326.

Miletitch, R., Reina, A., Dorigo, M., & Trianni, V. (2019). Emergent naming of resources in a foraging robot swarm. *arXiv preprint arXiv:1910.02274*.

Padilla-Iglesias, C., Gjesfjeld, E., & Vinicius, L. (2020). Geographical and social isolation drive the evolution of austronesian languages. *PloS one*, *15*(12), e0243171.

Steels, L. (2017). Human language is a culturally evolving system. *Psychonomic bulletin & review*, *24*(1), 190–193.

Steels, L., & Loetzsch, M. (2012). The grounded naming game. *Experiments in cultural language evolution*, *3*, 41–59.

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43.

Thomas, J., & Kirby, S. (2018). Self domestication and the evolution of language. *Biology & philosophy*, *33*(1), 1–30.

Trianni, V., De Simone, D., Reina, A., & Baronchelli, A. (2016). Emergence of consensus in a multi-robot network: from abstract models to empirical validation. *IEEE Robotics and Automation Letters*, *1*(1), 348–353.

# DECOUPLING SPEED OF CHANGE AND LONG-TERM PREFERENCE IN LANGUAGE EVOLUTION: INSIGHTS FROM ROMANCE VERB STEM ALTERNATIONS

Chundra A. Cathcart*[1,2], Borja Herce[1,2], and Balthasar Bickel[1,2]

*Corresponding Author: chundra.cathcart@uzh.ch
[1]Department of Comparative Language Science, University of Zurich, Switzerland
[2]Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

Romance verb stem alternations (e.g., Spanish ***teng**o* 'I have' vs. ***tien**es* 'you have') constitute seemingly unnecessary but highly inheritable morphological traits. Using novel phylogenetic methods, we assess the impact of frequency and alternation patterns on properties of their evolution, specifically on the speed of change and the long term preference for pattern types within lemmata. We find credible differences in long-term trends between alternation patterns, and confirm the notion that frequency drives the maintenance of irregular patterns. However, our model reveals no or only weak effects of either predictor on the speed of change. Our findings call for modeling the multiple dimensions of language change jointly but with distinct parameters for speed (or rates) of change and long-term (i.e., stationary) preferences.

## 1. Introduction

Stem alternations in Romance verb paradigms are well studied (Esher, 2016; Maiden, 2018; Herce, 2019, etc.). Their importance lies in that they constitute a prime example of unmotivated yet systematic patterns (Aronoff, 1994; Bickel, 1994). Unlike morphological alternations that express differences in meaning (e.g. English *drive* PRESENT vs. *drove* PAST) or are phonologically conditioned (e.g., *cat*[s] vs. *dog*[z]), Romance stem alternations do not mark a specific semantic value and are phonologically unpredictable. Despite this lack of motivation, these patterns exhibit remarkable diachronic stability, frequently spreading to new verbs. Here we seek to quantify the evolutionary dynamics of these patterns in an explicit phylogenetic framework, testing effects of frequency and alternation type.

The method we develop allows us to assess the effect of these two predictors on different components of language change, namely speed and stationary probability, interpretable as the long-term preference for a feature. Past research has either focused on speed of change (Pagel, Atkinson, & Meade, 2007; Greenhill et al., 2017) or on biases in change and preferred configurations (Maslova, 2000; Cysouw, 2011; Bickel, 2015; Jäger & Wahle, 2021). Little work to date has considered these properties of change jointly. We provide a principled means of teasing apart the dynamics of these two facets of language evolution, providing a more flexible understanding of the role of different factors in change.

## 2. Background

The philological literature identifies three main types of stem alternations in Romance verbs, labeled N, L, and P(YTA); these can co-occur within verbal paradigms. The emergence of the N and L patterns (see Table 1) occurred as a result of sound changes after Classical Latin but (largely) before the break-up of Romance into different languages. Thus, the alternation between *e* and *ie* in the stem of Spanish *perder* is due to different trajectories of change affecting stressed (leading to *ie*) and unstressed (leading to *e*) syllables. The alternation between *g* and *c* (phonetically [g] and [θ], respectively) in the stem of Spanish *decir* results from the palatalization of Latin *c* ([k]) before the vowels *e* and *i*.

Table 1. Partial paradigms of Spanish *perder* 'lose' and *decir* 'say'

| | N | | | L | | |
|---|---|---|---|---|---|---|
| | Pres. ind. | Pres. sbjv. | Ipf. ind. | Pres. ind. | Pres. sbjv. | Ipf. ind. |
| 1SG | pierd-o | pierd-a | perd-ía | dig-o | dig-a | dec-ía |
| 2SG | pierd-es | pierd-as | perd-ías | dic-es | dig-as | dec-ías |
| 3SG | pierd-e | pierd-a | perd-ía | dice | dig-a | dec-ía |
| 1PL | perd-emos | perd-amos | perd-íamos | dec-imos | dig-amos | dec-íamos |
| 2PL | perd-éis | perd-áis | perd-íais | dec-ís | dig-áis | dec-íais |
| 3PL | pierd-en | pierd-an | perd-ían | dic-en | dig-an | dec-ían |

While the alternations seen in these lexemes are expected developments due to regular sound change (e.g. *digo*, *di*[θ]*es* < *dī*[k]*ō*, *dī*[k]*is*), other lexemes have lost their alternations (e.g. *cue*[θ]*o*, *cue*[θ]*es* ← *coquō*, *coquis* 'cook' vs. expected †*cuego*, *cue*[θ]*es*), or have acquired the alternations in analogy with other verbs (e.g. *caigo*, *caes* ← *cadō*, *cadis* 'fall' vs. expected †*cao*, *caes*).

Unlike the other two patterns, the P pattern is inherited from Latin (see Table 2). In Latin, *fēc-* vs *fac-* marked a semantic distinction that is no longer present in modern Romance languages, leading to alternations that are arbitrary from the perspective of meaning. Like the other patterns, however, the P pattern has also been lost in some cases or expanded by analogy in others.

Irregular alternations like these have increasingly been studied in quantitative frameworks and cross-linguistically, exploring the roles of different predictors (e.g., frequency, word length, uniqueness points) in their maintenance (Marzi, Ferro, & Pirrelli, 2019; Sims-Williams, 2021). However, data-driven studies to date have not made use of a phylogenetic framework, which has the potential to shed light on evolutionary pressures that foster and militate against irregularity. In response, we develop a novel phylogenetic model to obtain a detailed understanding of the dynamics of change of these irregular patterns.

Table 2. Partial paradigms of Spanish *hacer*, and Latin *faciō* ’do’ illustrating P

| | Spanish | | | Latin | | |
|---|---|---|---|---|---|---|
| | Ipf. ind. | Pret. ind. | Ipf. sbjv. | Ipf. ind. | Perf. ind. | Plup. sbjv. |
| 1SG | hac-ía | hic-e | hic-iese | fac-iēbam | fēc-ī | fēc-issem |
| 2SG | hac-ías | hic-iste | hic-ieses | fac-iēbās | fēc-istī | fēc-issēs |
| 3SG | hac-ía | hi[θ]-o | hic-iese | fac-iēbat | fēc-it | fēc-isset |
| 1PL | hac-íamos | hic-imos | hic-iesemos | fac-iēbāmus | fēc-imus | fēc-issēmus |
| 2PL | hac-íais | hic-isteis | hic-ieseis | fac-iēbātis | fēc-istis | fēc-issētis |
| 3PL | hac-ían | hic-ieron | hic-iesen | fac-iēbant | fēc-ērunt | fēc-issent |

## 3. Data

In the Oxford Online Database of Romance Verb Morphology (Maiden, Smith, Cruschina, Hinzelin, & Goldbach, 2010; Beniamine, Maiden, & Round, 2020) each lemma is represented with its paradigm and is coded for its Latin etymon. For each lemma in each variety, we manually assessed whether it contained a reflex of the stem alternation patterns (N, L, and P) presented in Section 2. We gloss over language-specific idiosyncrasies in the inherited distribution of alternants in the paradigm for the purposes of this study (e.g., the L pattern affects the 3PL present indicative of Italian *dire* ‘say’ but is absent from the 1–2PL present subjunctive, in contrast to its Spanish cognate in Table 1), coding patterns as PRESENT/ABSENT/MISSING DATA. Among Latin-descended lemmata, 231 cognate sets are found in the database. We target items with wide coverage, excluding cognate sets that are limited to dialects of a single language or attested in fewer than five varieties. Additionally, we exclude uninformative lemma-pattern pairs that are completely present or absent across all languages with attested data. In total, we analyze 171 lemma-pattern pairs involving 66 lemmas from 67 varieties.

## 4. Method[1]

We investigate the historical dynamics of stem alternation using a phylogenetic comparative model. Methods of this sort require a timed phylogenetic representation of the languages in our data set in the form of a Bayesian tree sample, which we infer using RevBayes (Höhna et al., 2016) on the basis of both automatically generated lexical cognacy data (Jäger, 2018) and sound class data indicating which speech sounds are present in each variety (Heggarty et al., 2019). We impose uncontroversial clade constraints on the tree topology along with lower and upper bounds for each clade’s date calibration drawn from the literature on Romance languages (Hall, 1974). We use a a Birth-Death tree prior (Yang & Rannala, 1997) and a General Time-Reversible model of character evolution (Tavaré, 1986), along with a relaxed clock with log-normally distributed branch-level rate multipliers and gamma-distributed variation across 4 rate classes. We run 1,000,000

---

[1] Code available at `https://github.com/chundrac/JCoLE2022-morphomes`

iterations of Markov chain Monte Carlo over 4 chains, thinning the sample to 100 trees after discarding the first half as burn-in. We scale branch lengths so that one unit represents a millennium of change.

Under our model, each binary lemma-pattern pair evolves independently over the phylogeny of Romance according to a continuous-time Markov process. This process is parameterized by a gain and loss rate, or alternatively, the speed of change (irrespective of direction) and the stationary probability of feature presence, interpretable as the long-term preference for a given feature over a phylogeny (irrespective of the speed of change). The gain and loss rate of a lemma-pattern pair with index $d \in \{1, ..., D\}$ are $\pi_d s_d \rho$ and $(1 - \pi_d) s_d \rho$. Here, $s_d \rho$ represents the speed of change for feature $d$, $s_d$ being a multiplier of the global speed $\rho$. We place a Uniform$(0, 10)$ prior over $\rho$, preventing changes from happening more frequently than ten times per millennium. The parameter $\pi_d$ is the stationary probability of presence for the lemma-pattern pair in question.

While evolutionary parameters in phylogenetic comparative methods are usually estimated without predictors, we model them in a hierarchical distributional regression framework (Bürkner, 2018), allowing both the speed of change and stationary probability for each lemma-pattern pair to vary as a function of multiple predictors. We assume speed multipliers $\boldsymbol{s}$ and stationary probabilities $\boldsymbol{\pi}$ to be normally distributed, with a logit link to keep values within $(0, 1)$:

$$\text{logit } s_d \sim \text{Normal}(\alpha^s + \beta^{s,\text{LEMMA}}_{\text{LEMMAID}_d} + \beta^{s,\text{PATTERN}}_{\text{PATTERNID}_d}, \sigma^s) \tag{1}$$

$$\text{logit } \pi_d \sim \text{Normal}(\alpha^\pi + \beta^{\pi,\text{LEMMA}}_{\text{LEMMAID}_d} + \beta^{\pi,\text{PATTERN}}_{\text{PATTERNID}_d}, \sigma^\pi) \tag{2}$$

In each sampling statement, $\alpha$ denotes the intercept, $\beta^{\text{LEMMA}}$ represents the contribution of each lemma type, and $\beta^{\text{PATTERN}}$ represents the contribution of each alternation type. The standard deviations of these distributions represent the variance in speed and stationary probability that are not explained by the predictors included. We model the contribution of lemma type as a monotonic function of each lemma's frequency in Latin texts (Tombeur, 1998); this involves a combination of a parameter representing the effect of moving from the lowest to the highest frequency and a simplex parameter representing the effect of moving along the cline of frequency (Bürkner & Charpentier, 2020). We treat pattern type as two dummy-coded factors, comparing the levels L and P to N, respectively. We place Normal$(0, 1)$ priors over all model parameters in statements (1–2) with the exception of simplex parameters and standard deviations $\sigma$, which receive Dirichlet$(1, ..., 1)$ and HalfNormal$(0, 1)$ priors, respectively. Posterior distributions for parameters are inferred using the R package CmdStanR (Gabry & Češnovar, 2021). In line with the most conservative criteria for hypothesis evaluation, we infer decisive evidence for the effect of a predictor if the 95% credible interval (CI) of the corresponding parameter excludes zero, and strong evidence in cases where the 95% CI overlaps with zero but the 85% CI does not.

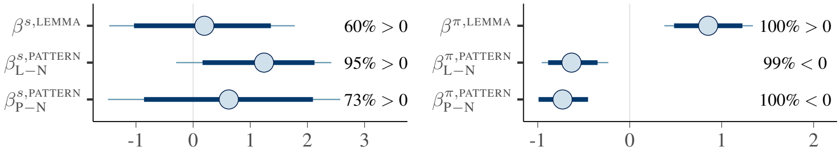


Figure 1. Medians and posterior 95% and 85% (shaded) credible intervals for model parameters of interest on the logit scale, along with percentage of samples above or below 0. PATTERN refers to the difference between N, L and P alternations; LEMMA refers to Latin lemma frequency.

## 5. Results

Posterior distributions for model parameters are given in Figure 1. In general, the predictors in our model do not appear to have a decisive effect on features' speed of change ($s$; left panel), though there is strong but not decisive evidence that the L pattern shows a higher speed of change than the N pattern. In contrast, we see decisive evidence for effects of all predictors on long-term pattern preference, i.e. the stationary probabilities $\pi$ (right panel). The 95% credible interval for the parameter $\beta^{\pi,\text{LEMMA}}$ is positive, indicating that more frequent lemmata are more likely to display a stem alternation pattern and less frequent lemmata are likelier to regularize. Additionally, we see decisive evidence that the long-term preference for N is greater than that of L and P, though post-hoc inspection of model parameters showed no evidence for a L vs. P contrast.

## 6. Discussion

Our results show that the N pattern outranks L and P in long-term preference (stationary probability) but not in stability (speed of change). A possible explanation might be that that the N pattern affects the 3SG present indicative, which tends to be highly frequent, whereas the L and P patterns do not.

More generally, we confirm the role of lexical frequency in the maintenance of irregular patterns. Arbitrary stem alternation patterns fly in the face of communicative and acquisitional efficiency in that they introduce more than one form per meaning. At the same time, irregularity enhances discriminability between forms in a paradigm, especially in more frequent lexemes (Nübling, 2011; Blevins, Milin, & Ramscar, 2017). Our results support the idea that more frequent lexemes are more likely to preserve and acquire irregular morphology over time.

The frequency effect is limited to stationary probabilties (long-term preferences), however, and does not affect how fast languages move to the preferred state (cf. $\beta^{\pi,\text{LEMMA}}$ vs $\beta^{s,\text{LEMMA}}$ in Fig. 1). This finding is consistent with the notion that cognitive and communicative pressures (such as frequency effects) — or indeed "language universals" more generally — bear primarily on stationary probabilities, and not on stability (Maslova, 2000; Cysouw, 2011; Bickel, 2015).

Our approach decouples these two aspects of evolution and facilitates explicit assessment and comparison of them in a single model. This flexibility is not generally found in recent work in evolutionary linguistics which tends to focus on variation in speed. We suspect that this emphasis stems from two sources: first, most prominent biological models of rate variation focus solely on speed variation (Huelsenbeck & Suchard, 2007; Heath, Holder, & Huelsenbeck, 2011), with few exceptions (e.g., Lartillot & Poujol, 2011). Second, a large body of work in language evolution revolves around change in basic vocabulary, where the replacement of one word by another cannot necessarily be construed as incurring a communicative cost in the same way that the development of unpredictable allomorphy might (though see Martin, 2007).

## 7. Conclusion

We investigated the development of irregular patterns in Romance verbal morphology using a novel method that unites phylogenetic modeling with hierarchical distributional regression modeling. Our results in some ways confirm received wisdom regarding the role of frequency in the maintenance of irregularity, but also shed light on poorly understood issues in language change, showing that frequency largely does not explain variation in speed of change but only in long-term preferences (stationary probabilities). One possible explanation for this is that speed of change responds to social pressure for differentiation such as schismogenesis, which is most strongly associated with vocabulary (Greenhill et al., 2017). Stem alternation and irregularity might be less accessible as markers of social differentiation and are therefore relatively immune to differences in speed of change. By contrast, they have direct implications for efficiency in processing, learning and communication, with effects on long-term preferences.

The model presented here can be expanded in a number of ways. It is straightforward to build in branch-level variation for speed and stationary probability, which can help to identify events of drastic change coinciding with language contact, schismogenesis, or other changes in the linguistic system. Additionally, our model used Latin lemma frequency as a proxy for an etymon's frequency throughout Romance history, a simplification that does not fully capture the dynamics of vocabulary change. In theory, it is possible to treat relative frequency as a continuous trait that varies over the tree (see Ringen, Martin, & Jaeggi, 2021 for flexible, complex models of the co-evolution of continuous and discrete cultural traits). A next step will involve building additional predictors and interaction terms into the model to investigate among other things whether phonologically and semantically similar lemmata undergo similar patterns of change, and whether or not other variables, such as conjugation class, play an interpretable role.

## Acknowledgements

This work was supported by the NCCR Evolving Language, Swiss National Science Foundation Agreement Nr. 51NF40_180888.

## References

Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes.* Cambridge, MA: MIT Press.

Beniamine, S., Maiden, M., & Round, E. (2020). Opening the Romance Verbal Inflection Dataset 2.0: a CLDF lexicon. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 3027–3035).

Bickel, B. (1994). In the vestibule of meaning: Transitivity inversion as a morphological phenomenon. *Studies in Language*, *19*, 73–127.

Bickel, B. (2015). Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis, 2nd edition* (pp. 901–923). Oxford: Oxford University Press.

Blevins, J. P., Milin, P., & Ramscar, M. (2017). The Zipfian paradigm cell filling problem. In *Perspectives on morphological organization* (pp. 139–158). Brill.

Bürkner, P.-C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, *73*(3), 420–451.

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411.

Cysouw, M. (2011). Understanding transition probabilities. *Linguistic Typology*, *15*, 415–431.

Esher, L. (2016). Morphomes and predictability in the history of Romance perfects. *Diachronica*, *32*(4), 494–529.

Gabry, J., & Češnovar, R. (2021). *Cmdstanr: R interface to 'CmdStan'.* (https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org)

Greenhill, S. J., Wu, C.-H., Hua, X., Dunn, M., Levinson, S. C., & Gray, R. D. (2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, *114*(42), E8822–E8829.

Hall, R. A. (1974). *Comparative Romance grammar.* New York/London/ Amsterdam: Elsevier.

Heath, T., Holder, M., & Huelsenbeck, J. (2011). A Dirichlet Process Prior for estimating lineage-specific substitution rates. *Molecular biology and evolution*, *29*, 939-55.

Heggarty, P., Shimelman, A., Abete, G., Anderson, C., Sadowsky, S., Paschen, L., Maguire, W., Jocz, L., Aninao, M. J., Wägerle, L., Dërmaku-Appelganz, D., Silva, A. P. d. C. e., Lawyer, L. C., Michalsky, J., Cabral, A. S. A. C., Walworth, M., Koile, E., Runge, J., & Bibiko, H.-J. (2019). *Sound Comparisons: Exploring Diversity in Phonetics across Language Families.*

Herce, B. (2019). Morphome interactions. *Morphology*, *29*(1).

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore,

B. R., Huelsenbeck, J. P., & Ronquist, F. (2016). Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, *65*(4), 726–736.

Huelsenbeck, J. P., & Suchard, M. A. (2007). A nonparametric method for accommodating and testing across-site rate variation. *Systematic Biology*, *56*(6), 975–987.

Jäger, G. (2018). Global-scale phylogenetic linguistic inference from lexical resources. *Scientific Reports*, *5*.

Jäger, G., & Wahle, J. (2021). Phylogenetic typology. *Front. Psychol.*, *12*.

Lartillot, N., & Poujol, R. (2011). A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Molecular biology and evolution*, *28*(1), 729–744.

Maiden, M. (2018). *The Romance verb: Morphomic structure and diachrony.* Oxford University Press.

Maiden, M., Smith, J. C., Cruschina, S., Hinzelin, M.-O., & Goldbach, M. (2010). *Oxford Online Database of Romance Verb Morphology.* University of Oxford. Online at http://romverbmorph.clp.ox.ac.uk/.

Martin, A. T. (2007). *The evolving lexicon.* Unpublished doctoral dissertation, University of California, Los Angeles.

Marzi, C., Ferro, M., & Pirrelli, V. (2019). A processing-oriented investigation of inflectional complexity. *Front. Comm.*, *4*, 1–23.

Maslova, E. (2000). A dynamic approach to the verification of distributional universals. *Linguistic Typology*, *4*, 307-333.

Nübling, D. (2011). How do exceptions arise? On different paths to morphological irregularity. In H. Simon & H. Wiese (Eds.), *Expecting the unexpected: exceptions in grammar* (p. 139-162). Berlin: de Gruyter.

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*, 717-720.

Ringen, E., Martin, J. S., & Jaeggi, A. (2021). Novel phylogenetic methods reveal that resource-use intensification drives the evolution of "complex" societies. *EcoEvoRxiv.*

Sims-Williams, H. (2021). Token frequency as a determinant of morphological change. *Journal of Linguistics*, 1–37.

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, *17*(2), 57–86.

Tombeur, P. (1998). *Thesaurus formarum totius Latinitatis a Plauto usque ad saeculum XXum.* Turnhout: Brepols.

Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular biology and evolution*, *14*(7), 717–724.

# SIMULATION OF EMERGENT COMMUNICATION WITH LARGE SCALE MACHINE LEARNING

Rahma Chaabouni*[1], Florian Strub*[1], Florent Altché[1], Corentin Tallec[1], Eugene Trassov[1], Elnaz Davoodi[1], Kory Mathewson[1], Olivier Tieleman[1], Angeliki Lazaridou[1], and Bilal Piot[1]*

*Corresponding Authors: {rahmac,fstrub,piot}@deepmind.com
[1]DeepMind

The origins of human languages are unarguably a complex research problem (Christiansen & Kirby, 2003). Over the last fifty years, this inherent complexity has led to various research approaches from many fields ranging from biology, linguistics, anthropology, computer science etc. (Harnad et al., 1976; Bickerton, 2007; Nölle et al., 2020). Here, we walk in the footprint of the computational community (Steels, 1997; Kirby, 2002; Myers-Scotton, 2002). Roughly, the computational approach aims to simulate the prerequisite or processes that could trigger the emergence of a structured language within a controlled environment. In practice, the classic modeling paradigm simulates two agents that must develop a communication protocol to solve the Lewis game; the first agent describes an object such as the second agent can locate it among a set of candidate objects. Therefore, the two agents must settle on a communication protocol to solve the game. In this paper, we discuss how new computational resources and recent machine learning algorithms, namely deep learning (LeCun et al., 2015) and reinforcement learning (Sutton & Barto, 2018), allow scaling language simulation emergence. We then discuss how this more realistic model can foster new research directions in language evolution.

Despite the recent astonishing successes of deep neural networks in solving complex tasks, computational emergent language has made little profit from these advances. Most works still consider small disentangled input spaces (Lazaridou & Baroni, 2020), where the expected language can often be reduced to a basic identity operator. In fact, modeling scenarios barely evolved over twenty years, e.g. Kirby (2002) and Ren et al. (2019) both use the same binary input vector of size eight, and only a few papers went beyond artificial input spaces (Havrylov & Titov, 2017; Lu et al., 2020). In this paper, we endorse Bickerton (2015)'s view about the necessity of complex tasks and stimuli to model human language communication. Furthermore, such challenging settings have been proven to be paramount to emulate complex distributions in the machine learning literature (Krizhevsky et al., 2012; Brown et al., 2020). We argue for enhancing the

Lewis Game to more realistic settings to gain novel and more conclusive insights.

We cast the language emergence modeling into a deep reinforcement learning framework similar to Lazaridou, Peysakhovich, and Baroni (2017). Then, we summarize the modeling tools and algorithms, and provide guidance towards good practices from the machine learning field[1]. As a core contribution, we focus on three independent aspects to challenge language emergence modeling: increasing task complexity, using complex visual inputs, generating large populations of agents. We then assess the different properties of emergent languages through; (1) Generalization by computing the agents' communication success at test time, (2) Topographic similarity as a proxy for compositionality (Brighton & Kirby, 2006), (3) Ease of learning by using the emergent protocol for transfer tasks.

First, we consider more complex tasks by increasing the number of candidates among which the listener must retrieve the target input. Specifically, agents typically discriminate between less than 20 candidates in the emergent language field (Mu & Goodman, 2021); we experiment with up to 1024 candidates in this study. We found that scaling up the task complexity entails unstable optimization. We propose to smooth language learning by using classic mathematical regularization. Furthermore, we observe that complexifying the task has two positive aspects: it better discriminates the different models and improves the generalization of the learned communication protocol.

Second, most emergent language studies situated agents in a simple one-hot vector environment (Ren et al., 2019; Rita et al., 2020), we here scale-up the input space with continuous and ambiguous visual cues by using pretrained representations of natural images (Grill et al., 2020). We note no correlation between generalization and the widely used topographic similarity metric in this set of experiments. We hence question if this metric is adequate to assess compositionality in complex setups. Inspired by the computer vision community (Grill et al., 2020), we then discuss how ease of learning may be surrogate for protocol evaluation.

Finally, we investigate the impact of population size. In particular, we scale up the Lewis game from 2 to up to 100 agents. Here, unlike what was observed in human communication (Gary Lupyan, 2010; Raviv et al., 2019), we find little to no systematic benefit on emergent languages' properties when increasing the population size. We propose alternative methods to leverage populations, namely voting and imitation among speakers (Hester et al., 2018). Our results show that such population dynamics lead to more robust, productive, and in some cases easy-to-learn languages, opening up new research opportunities.

In the end, we expect that these observations, baselines, and good practices would allow the language emergence community to benefit further from deep RL advances. We believe that such a more realistic and challenging framework is a prerequisite to moving the field closer to its goals of modeling language evolution.

---

[1] Code: `https://github.com/deepmind/emergent_communication_at_scale`

## References

Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua*, *117*(3), 510–526.

Bickerton, D. (2015). *Roots of language.* Language Science Press.

Brighton, H., & Kirby, S. (2006). Understanding linguistic evolution by visualizing the emergence of topographic mappings. *Artificial life*, *12*(2), 229–242.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al.. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Christiansen, M. H., & Kirby, S. E. (2003). *Language evolution.* Oxford University Press.

Gary Lupyan, R. D. (2010). Language structure is partly determined by social structure. *PLoS ONE 5*, *1*.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., et al.. (2020). Bootstrap your own latent-a new approach to self-supervised learning. *Proc. of NeurIPS*.

Harnad, S., Steklis, H. D., & Lancaster, J. B. (1976). Origins and evolution of language and speech.

Havrylov, S., & Titov, I. (2017). Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In *Proc. of NeurIPS.*

Hester, T., Vecerik, M., Pietquin, O., Lanctot, M., Schaul, T., Piot, B., Horgan, D., et al.. (2018). Deep q-learning from demonstrations. In *Proc. of AAAI.*

Kirby, S. (2002). Natural language from artificial life. *Artificial life*, *8*(2), 185–215.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Proc. of NeurIPS.*

Lazaridou, A., & Baroni, M. (2020). Emergent multi-agent communication in the deep learning era. *arXiv preprint arXiv:2006.02419*.

Lazaridou, A., Peysakhovich, A., & Baroni, M. (2017). Multi-agent cooperation and the emergence of (natural) language. In *Proc. of ICLR.*

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Lu, Y., Singhal, S., Strub, F., Courville, A., & Pietquin, O. (2020). Countering language drift with seeded iterated learning. In *Procof ICML.*

Mu, J., & Goodman, N. (2021). Emergent communication of generalizations. In *Proc. of NeurIPS.*

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes.* Oxford, UK: Oxford University Press.

Nölle, J., Hartmann, S., & Tinits, P. (2020). Language evolution research in the year 2020: A survey of new directions. *Language Dynamics and Change*.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*(1907), 20191262.

Ren, Y., Guo, S., Havrylov, S., Cohen, S., & Kirby, S. (2019). Enhance the compositionality of emergent language by iterated learning. In *Proc. of EmeCom - Neurips).*

Rita, M., Chaabouni, R., & Dupoux, E. (2020). "LazImpa": Lazy and impatient neural agents learn to communicate efficiently. In *Proc. of ACL.*

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of communication*, *1*(1), 1–34.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press.

# LINGUISTIC DENSITY IN A GLOBAL SCALE IS DETERMINED BY SOCIAL, POLITICAL AND ECONOMIC FACTORS

THIAGO CHACON*[1], MARC ALLASSONNIERE-TANG[2]

*Corresponding Author: name@domain.com
[1]University of Brasilia / IEA-Paris, Brazil / France
[2]Muséum National d'Histoire Naturelle, Paris, France

The worldwide distribution of languages and linguistic families is uneven, with some areas presenting higher language and phylogenetic densities than others. Previous research has explored the role of environmental, social and economic factors in explaining global patterns of language and phylogenetic densities. For example, Nichols (1997: 383) claims that “linguistic density is highest in areas where small societies can be more or less autonomous on small territories”, noticing a correlation of environmental and political-economic factors in a way that “densities are higher in coastal regions, at lower latitudes, and in wetter and less seasonal climates” (op.cit.). Nettle (1998) attributes a causal effect from climate (year round mean growing season) on social networks and its direct impact on language areas: “where the climate allows continuous food production throughout the year, small groups of people can be reliably self-sufficient and so populations fragment into many small languages”. Hua et al. (2019) build on Nettle’s hypothesis and show that temperature and precipitation seasonality are specific climatic factors associated with language densities, while showing that climate has no direct impact on the average population size of social groups. Currie and Mace (2009) have shown that high political centralization of a society is a factor determining the size of language areas, in a way that languages whose societies have more centralized political organization will be spoken over larger areas, which ultimately diminish the linguistic density within the areas where these languages are spoken.

Our study confirms and makes further advancements from the latter ones by investigating the effects of social, political, economic and environmental variables in determining the global distribution of areas with high concentration of languages and linguistic families. For social variables, we use D-Place (Kirby et al. 2016) to test parameters related to marriage organization (EA015), dependence on agriculture (EA05), political complexity (EA033) and ethnic descent (EA043), which are found across 818 societies worldwide. Population size is also included. The information of language richness (at the language and at the family level) and the information of ecological factors (precipitation rate in wettest quarter and roughness of terrain) were extracted based on the method of Derungs et al (2018), where the world map is cut into grids (we tested resolutions from 295 to 3267 point-grids) and each society found within a given grid is annotated with the value of that grid for the environmental variables.

Our results based on linear mixed models controlling for language family and geographic area as random effects (and considering interaction between the variables) show that political autonomy, high population densities, the existence of ethnic boundaries such as clans and greater reliance on agriculture are positively associated with areas with high linguistic densities. As the concentration of distinct languages and linguistic families in an area is ultimately dependent on migration, diversification, and the long-term sustainability of local diversity, these results suggest that social, political and economic factors are relevant in creating greater autonomy and boundaries that act as triggers of language splits and/or as buffers to language shift.

As the high concentration of distinct languages and linguistic families in an area is ultimately dependent on migration, in-situ diversification events, and the long-term sustainability of local diversity, these results suggest that social, political and economic factors are relevant in creating greater autonomy and boundaries that act as triggers of language splits and as buffers to language shift. Politically independent societies can avoid dominance by external groups, can more often cause splits within a social group as well as use language boundaries to express differences in identity from neighboring groups. Endogamous societies can reproduce the ethnolinguistic group with less dependence on their neighbors, which gives them greater autonomy to migrate and split-off from related groups, as well as diminishes the effects of demographic pressure from speakers of distinct languages, which could cause the linguistic homogenization of an area. Agriculture can be seen not as a factor causing language spread and shift, but as an economic strategy that ensures greater autonomy of local groups. As we face an increasing threat to linguistic diversity on a global scale, this study suggests that greater autonomy for indigenous self-governance strategies can be the key to the sustainability of the world's linguistic and cultural diversity.

## References

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2021). Global predictors of language endangerment and the future of linguistic diversity. Nature Ecology & Evolution. https://doi.org/10.1038/s41559-021-01604-y

Currie, T. E., & Mace, R. (2009). Political complexity predicts the spread of ethnolinguistic groups. *Proceedings of the National Academy of Sciences*, 106(18), 7339–7344. https://doi.org/10.1073/pnas.0804698106

Derungs, C., Köhl, M., Weibel, R., & Bickel, B. (2018). Environmental factors drive language density more in food-producing than in hunter–gatherer populations. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2017.2851

Hua, X., Greenhill, S. J., Cardillo, M., Schneemann, H., & Bromham, L. (2019). The ecological drivers of variation in global language diversity. Nature Communications, 10(1), 2047. https://doi.org/10.1038/s41467-019-09842-2

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., Leehr, D., Low, B. S., McCarter, J., Divale, W., & Gavin, M. C. (2016). D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *PLOS ONE*, 11(7), e0158391. https://doi.org/10.1371/journal.pone.0158391

Nettle, D. (1998). Explaining Global Patterns of Language Diversity. *Journal of Anthropological Archaeology*, 17(4), 354–374. https://doi.org/10.1006/jaar.1998.0328

Nichols, J. (1997). Modeling Ancient Population Structures and Movement in Linguistics. *Annual Review of Anthropology*, 26(1), 359–384. https://doi.org/10.1146/annurev.anthro.26.1.359

# A COMPUTATIONAL SIMULATION OF THE GENESIS AND SPREAD OF LEXICAL ITEMS IN SITUATIONS OF ABRUPT LANGUAGE CONTACT

Ching-Yat Cheung *[1,2], Kofi Yakpo [1], Christophe Coupé [1,3]

*Corresponding Author: chingyat@connect.hku.hk
[1]Department of Linguistics, The University of Hong Kong, Hong Kong SAR
[2] LEADS Group, Max Planck Institute for Psycholinguistics, The Netherlands
[3] Laboratoire Dynamique Du Langage, Centre National de la Recherche Scientifique, France

The current study presents an agent-based model which simulates the innovation and competition among lexical items in cases of language contact. It is inspired by relatively recent historical cases in which the linguistic ecology and sociohistorical context are highly complex. Pidgin and creole genesis offers an opportunity to obtain linguistic facts, social dynamics, and historical demography in a highly segregated society. This provides a solid ground for researching the interaction of populations with different pre-existing language systems, and how different factors contribute to the genesis of the lexicon of a newly generated mixed language. We take into consideration the population dynamics and structures, as well as a distribution of word frequencies related to language use, in order to study how social factors may affect the developmental trajectory of languages. Focusing on the case of Sranan in Suriname, our study shows that it is possible to account for the composition of its core lexicon in relation to different social groups, contact patterns, and large population movements.

## 1. Introduction

A consequence of colonialism and population movements is the genesis of a multilingual and sociologically complex ecology, in which people of diverse language backgrounds are brought into contact, creating the need for a new shared means for communication. In some of these cases, a pidgin or a creole, which is characterized by the mixture of lexical and grammatical features from different languages in the language ecology, emerges to answer communicative needs. Such recent emergence of languages may shed light on the overall evolutionary process of human languages (Mufwene, 2008). Specifically, population movements have been argued as one of the factors that motivate the evolution of contact languages (Mufwene, 2007).

The present study presents a computational model which serves as a complement to empirical studies when investigating how population movements and social structures affect the emergence of a language and specifically the developmental trajectory of its lexical inventory. The paper expands on previous studies by considering social structures, population movements, communicative needs, and the functional frequency of the lexical items. Our model is specifically inspired by the historical case of Suriname, a former colony in South America that was ruled by English and Dutch speakers. The Dutch took over the colony from the English, which led to an exodus of English native speakers and influenced the development of Sranan, an English-lexified creole spoken in the area (Arends, 2017). Like other contact languages, it is characterized by a mixture of various languages spoken in the language ecology. We thus constructed a model to investigate how the sudden departure of a dominant social group may have an influence on the collective convergence towards a linguistic form when there is a language competition. Essentially, the agent-based model allows us to experiment with several sociolinguistic variables and observe different population movements can affect language evolution. As reported below, the model demonstrates the possibility for agents to develop a multilingual - rather than monolingual - lexical inventory. The predictions also align quantitatively with the lexical data of Sranan.

## 2. Previous Studies

There have been recent attempts to construct computational models to better understand contact languages. Jansson et al. (2015), inspired by the Naming Game (Baronchelli et al., 2006), simulated the evolution of the lexicon, phonology and syntax of Mauritian Creole using data on the demography and population movements in Mauritius in the 18th century. Specifically for the lexicon, their model showed how agents can converge to one single lexifier (i.e., French) solely based on communicative needs. Tria et al., (2015) also employed a computational model to study the possible emergence of creoles. Their model – a modified version of the Naming Game with non-trivial interaction rules based on the observation of communicative pressures in highly segregated societies – revealed that such an emergence can be accurately predicted on the sole basis of historical demographic data, more precisely the proportion of different social/ethnic groups in the multilingual ecology. Additionally, Furman & Nitschke (2020) investigated the role of external factors which affect the evolution of creole languages, such as the population size of different interacting groups and the lexical similarity between these groups, in the convergence of lexical items with an iterative agent-based naming game.

Recently, Cheung (2022) further showed that computational social networks with different density and connectivity patterns within the same social group and across different social groups – which can differ due to their economic structures (such as ‘sparse’ cotton plantations and dense urban areas/ports) – could affect the likelihood and rate of creole emergence. These results, which are specific to scenarios of recent abrupt language contact, can be related to more generic studies that attempt to model language competition, change and death in a multilingual society (e.g. Abrams & Strogatz, 2003; Minett & Wang, 2008), especially when different social structures (Ke et al., 2008; Loureiro-Porto & Miguel, 2017) are found to be relevant.

The previous attempts have all inspired the present study.

## 3. Modelling the lexical propagation of historical creoles

Our model is inspired by the historical case of Suriname, a former colony in South America. Historical data (Migge, 1998, based on Postma, 1990 and Voorhoeve & Lichtveld, 1975) show the demographic evolution of each social group/ethnicity of the region through time. There was a significant drop of the English-speaking population from 1668 due to the colony being ceded by the British Empire to the Netherlands under the Treaty of Breda in 1667 (Arends, 2017). There were approximately 2000 English speakers in 1666, but the number dropped to 820 in 1668, and there were only 38 still remained in 1680 (Voorhoeve & Lichtveld, 1975). The number of other non-English Europeans, mainly Dutch, increased slightly, while the number of Blacks also increased. The exodus of English native speakers reportedly influenced the development of Sranan, an English-lexified creole spoken in the area (Arends, 2017).

Our computational model employs an agent-based naming game that simulates repeated interactions between agents. Each agent initially has their own lexicon of 300 basic words in their own language. Each lexical referent is assigned with a probability drawn from the Zipfian distribution, which indicates the chance of being picked out and used during an interaction, such that we can partially account for language use.

Agents can be assigned with one of three fixed roles: Blacks, European group 1 (E1) (the English) or European group 2 (E2) (the Dutch). The population of Blacks is further divided into five groups of rural plantation workers (R1-R5) and an urban group (U). Members of each rural group are sparsely distributed and related according to a scale-free social network structure (the average degree of each node is ~1.93), while each of the European and Black Urban groups is a dense small-world network (rewiring probability of 0.05 with an average degree of 8). Random links across the different groups are added in proportion to the total population of these groups. This hypothetical network setup is based on other studies of models and empirical cases (c.f. Cheung, 2022).

Given the situation, an agent may know up to four words for a given referent: one in English (E1), one in Dutch (E2), one in an African language (A), and one in an emerging creole (C). We modify the interaction rules presented in Tria et al. (2015), where all the agents could interact with each other. In our model, for each interaction, an existing edge between two agents is randomly picked and these agents play a naming game with a random assignation of the roles of speaker and listener. If the communication is successful (i.e., the listener knows the speaker's word , in the language they chose, for the chosen referent), any other word – in another language – for the referent is discarded. If the communication fails, the listener picks up the word from the speaker. While this widely used paradigm may not be the most accurate reflection of the underlying cognitive mechanisms of word learning, it offers a plausible model for the switch of belief between different lexical representations for a particular referent in order to choose which one is most suitable to foster communication success.

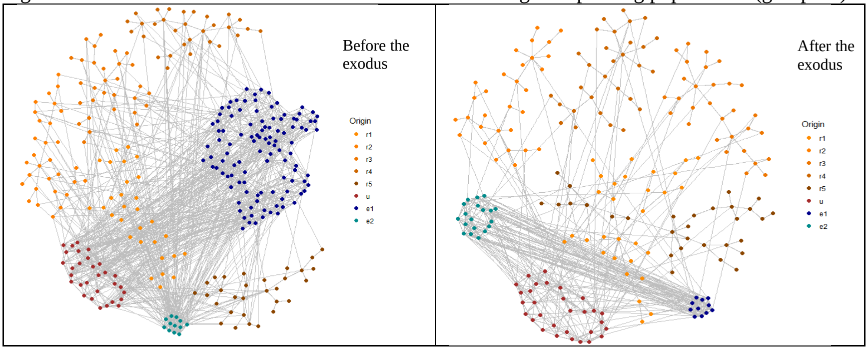
The interaction is governed by several constraints reflecting the segregated social dynamics. Even if both speaker and listener have a word in an African language for a particular referent, they will only be able to communicate successfully with a probability $\delta$. This parameter reflects the actual diversity of African languages spoken by Blacks. The Europeans (*E1, E2*) only have a probability $\beta$ to learn a word in an African language successfully (this includes European-to-European communication), and they further only have a probability $\varepsilon$ to learn from the Blacks (*R/U*), due to differences in socioeconomic status. The chance of creole innovation, if the listener possesses both a European and an African language for a lexical referent, is given by $\gamma$, and the chance of a listener adopting this creole is given by $\alpha$, an adjustment made to represent the failure of learning due to potential difficulties in interpreting the mixing and the preference to adapt items from the superstratum. Given the case of Suriname and our focus on the effect of population movements, we assumed a reasonable fixed value for each of the parameters across all conditions: $\alpha = 0.8$, $\beta = 0.2$, $\delta = 0.25$, $\varepsilon = 0.25$ and $\gamma = 0.8$ (the parameters could of course be modified to explore significantly different language ecologies)

In order to investigate the consequences of the sudden exit of the English speakers, a large proportion (over 60%) of the members of group 1 (E1) is removed after a number of interactions/steps – with two possible departure times (Conditions 1 and 2). Figure 1 illustrates how such an exodus transforms the overall social network. Furthermore, we consider the condition in which the graph remains unchanged throughout the whole simulation as a baseline (Condition 3). We finally model a condition (Condition 4), a setup based on condition 1, in which both groups of Europeans are removed after a number of interactions to simulate the development of a creole among the Blacks themselves – a situation typical of creole development. Table 1 illustrates the 4 conditions of the simulations.

Table 1: the different conditions that were used in the simulations

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | E1 departs<br>E2 increases | | | | | | End |
| 2 | | | E1 departs<br>E2 increases | | | | End |
| 3 | | | | | | | End |
| 4 | E1 departs<br>E2 increases | | | | Both E1 &<br>E2 depart | | End |
| Steps | 100,000 | … | 1,000,000 | ……… | 2,000,000 | …………………… | 5,000,000 |

Figure 1: an illustration of the simulation of an exodus of English-speaking populations (group E1)



## 3. Results

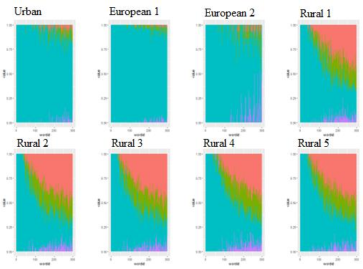
The resulting lexicons are assessed for the different social groups. We find that the departure of the native speakers of European languages affects the final stage of the process of linguistic convergence. The speakers remain under communicative pressure due to multilingualism after the exit of the speakers of the superstrate languages, and they still need to develop a lingua franca, causing more creole innovation and language diversification. This is in alignment with previous observations in the literature (Mufwene, 1996, 2007).

Figure 2 shows the results after 5 million steps of interaction in our different conditions. Across all conditions, the agents are under communicative pressure and must converge to communicate with each other. An observation that can be made from the simulation under Condition 3 is that if the native speakers of the main lexifier, English, remain throughout the simulation, all agents tend to converge to this single language, due to sustained linguistic input and sufficient exposure. Furthermore, in condition 2, in which English speakers depart relatively late (at the 1-millionth step), it can be observed that although a majority of the words still stem from their language, some mid-frequency words are adopted from the other European language, Dutch. For both conditions 2 and 3, no significant amount of creole innovation occurs (i.e., no convergence to creole words). In condition 1, the native speakers of the main language exit early, and while their language still occupies a large proportion of the final lexicon, Dutch is also used to denote some words that are of higher frequency.

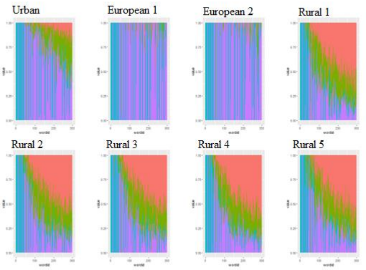
In figure 3, a comparison of Conditions 1 and 4 shows that creole words are more prominent in the latter condition. Condition 4, in particular, appears to be the closest approximation of the etymological distribution of Sranan, in which it is estimated that there are roughly 18% English words, 21.5% Dutch words, and 36% innovations (Romaine, 2001).

Figure 2: Etymological composition of the lexicon for different conditions after 5 million steps

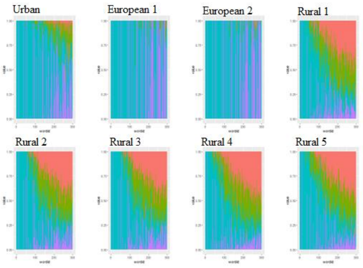
Condition1: E1 departs early at the 100k$^{th}$ step



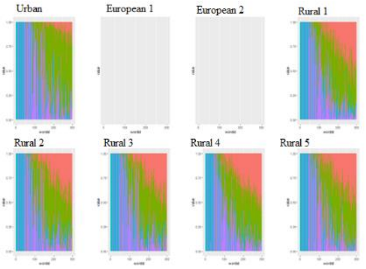
Condition 2: E1 departs late at the 1m$^{th}$ step



Condition3: No group departs throughout

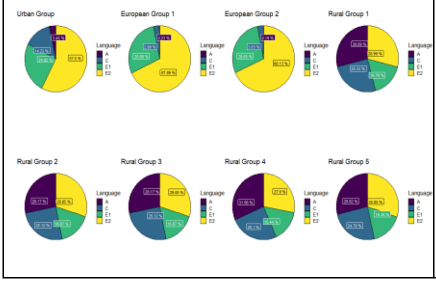
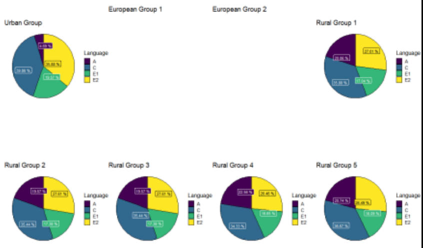


Condition 4: E1 departs at the 100k$^{th}$ step, E1 & E2 depart at the 2-millionth step



The x-axis corresponds to the lexical items ranked by decreasing frequency from left to right, and the y-axis reports the average proportion of each language across all agents (Blue: E1; Purple: E2; Green: C; Red: A). Each plot represents the situation of a social group.

Figure 3: cumulative composition of the lexicon across different groups (Purple: A, Blue: C, Green: E1, Yellow: E2)

| Condition 1: E1 departs early at the 100kth step | Condition 4: E1 departs at the 100kth step, E1 & E2 departs at the 2-millionth step |
|---|---|
|  |  |

In addition, due to the high segregation of the society, there is a delay in the rural area in terms of linguistic exposure. This suggests a possible variation continuum between the urban and the rural areas, which is also observed in Suriname (Arends, 2017). The study of the interaction between language variety and social stratification could be an interesting direction for further exploration.

## 4. Conclusion

The present model offers a plausible way to computationally investigate socio-historically complicated and empirically specific scenarios of language contact. It offers reasonable predictions for a given social dynamics and segregated relationships between different social groups, and the group selection of lexical items appears to be highly sensitive to the interactions and language use between agents.

The current model could be extended to other (socio)linguistic phenomena, such as the social dynamics during second language learning. The observation of the different intermediate stages of language development could also explain how a variation continuum could appear between various social groups. Meanwhile, the model mostly focuses on horizontal transmission of language, and the role of vertical transmission is underplayed. It would be worthwhile to model the birth of children with initially blank-slate agents (in terms of linguistic repertoire and exposure), and observe how this affects the linguistic dynamic.

## References

Abrams, D. M., & Strogatz, S. H. (2003). Modelling the dynamics of language death. *Nature*, *424*(6951), 900–900. https://doi.org/10.1038/424900a

Arends, J. (2017). *Language and Slavery: A social and linguistic history of the Suriname creoles* (Vol. 52). John Benjamins Publishing Company. https://doi.org/10.1075/cll.52

Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., & Steels, L. (2006). Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment*, *2006*(06), P06014–P06014. https://doi.org/10.1088/1742-5468/2006/06/P06014

Cheung, C. Y. (2022). *Investigating contact languages with agent-based modelling*. The University of Hong Kong.

Furman, G., & Nitschke, G. (2020). Evolving an Artificial Creole. *Proceedings of The Genetic and Evolutionary Computation Conference 2020 (GECCO '20).*, 2.

Jansson, F., Parkvall, M., & Strimling, P. (2015). Modeling the Evolution of Creoles. *Language Dynamics and Change*, *5*(1), 1–51. https://doi.org/10.1163/22105832-00501005

Ke, J., Gong, T., & Wang, W. S.-Y. (2008). Language Change and Social Networks. *Commun. Comput. Phys.*, 15.

Loureiro-Porto, L., & Miguel, M. S. (2017). Language Choice in a Multilingual Society: A View from Complexity Science. In S. S. Mufwene, C. Coupe, & F. Pellegrino (Eds.), *Complexity in Language* (pp. 187–217). Cambridge University Press. https://doi.org/10.1017/9781107294264.008

Migge, B. (1998). Substrate influence in creole formation: The origin of give-type serial verb constructions in the Surinamese Plantation Creole. *Journal of Pidgin and Creole Languages*, *13*(2), 215–266. https://doi.org/10.1075/jpcl.13.2.02mig

Minett, J. W., & Wang, W. S.-Y. (2008). Modelling endangered languages: The effects of bilingualism and social structure. *Lingua*, *118*(1), 19–45. https://doi.org/10.1016/j.lingua.2007.04.001

Mufwene, S. S. (1996). The Founder Principle in Creole Genesis. *Diachronica, 13*(1), 83–134. https://doi.org/10.1075/dia.13.1.05muf

Mufwene, S. S. (2007). Population Movements and Contacts in Language Evolution. *Journal of Language Contact*, *1*(1), 63–92. https://doi.org/10.1163/000000007792548332

Mufwene, S. S. (2008). *Language evolution: Contact, competition and change*. Continuum.

Postma, J. (1990). *The Dutch in the Atlantic Slave Trade, 1600–1815* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511528958

Romaine, S. (2001). Lexical structure in Pidgins and Creoles. In D. A. Cruse, F. Hundsnurscher, M. Job, & P. R. Lutzeier (Eds.), *Lexikologie* (pp. 1092–1095). De Gruyter. https://doi.org/10.1515/9783110171471.2.26.1092

Tria, F., Servedio, V. D. P., Mufwene, S. S., & Loreto, V. (2015). Modeling the Emergence of Contact Languages. *PLOS ONE*, *10*(4), e0120771. https://doi.org/10.1371/journal.pone.0120771

Voorhoeve, J., & Lichtveld, U. M. (Eds.). (1975). *Creole drum: An anthology of Creole literature in Surinam*. Yale University Press.

# USING EYE TRACKING TO BETTER UNDERSTAND CHILDREN'S PROCESSING OF EVENTS DURING VERB LEARNING: IS THE FOCUS ON PEOPLE (FACES) OR THEIR ACTIONS (HANDS)?

JANE B. CHILDERS*[1], EMILY WARKENTIN[1], MARISSA YOUNG[1], SNEH LALANI[1], and AKILA GOPALKRISHNAN[1]

*Corresponding Author: jchilder@trinity.edu
[1]Psychology Department, Trinity University, San Antonio, Texas USA

Verbs are central to the syntactic structure of sentences, and thus important for learning one's native language. This study examines how children are visually inspecting events as they hear, and do not hear, a new verb. Specifically, there is evidence that children may focus on the agent of the action (Maguire et al., 2008; Childers et al., 2017) or may prioritize attention to the action being performed (e.g., Behrend, 1990); to date, little evidence is available. This study uses an eye tracker to track participants' looking to the agent (i.e., face) vs. action (i.e., hands) when viewing events linked to a new verb as well as distractor events to better understand children's visual attention patterns. Two 1/2, 3 1/2- and 4 1/2-year-olds saw dynamic target and distractor events and heard new verbs while a Tobii x30 eye tracker recorded fixations to AOIs (head/face, hands). Results show that children in all age groups were able to learn and extend the new verbs to new events at test. Additionally, across age groups, when viewing target events, children increased their looking to the hands (where the action is taking place) as those trials progressed and decreased their looking to the agents' face, which is less informative for learning a new verb's meaning. In contrast, when viewing distractor events, children decreased their looking to hands over trials and maintained their attention to the face. In sum, children's visual attention to agents' faces and hands differed depending on whether the events cooccurred with the new verb. These results are important as this is the first study to show this pattern of visual attention during verb learning, and thus these results help reveal underlying attentional strategies children may use when learning verbs. Additionally, the study of children's acquisition of verbs, which vary across languages, could provide clues to how languages may have changed in response to the learnability of emerging predicate structures.

## 1. Introduction

Verbs are central to the syntactic structure of sentences. A controversy in this area is whether children focus on the agent of the action or the action being performed, and to what extent they focus on each one. Some studies suggest that young children learn verbs best when a single agent is seen, which could mean that they are attending too much to agents rather than the actions while learning verbs (Maguire, Hirsh-Pasek, Golinkoff & Brandone, 2008; Childers, Paik, Flores, Lai & Dolan, 2017: complex events). In other studies, children seem to be attending to actions (e.g., Scott & Fisher, 2012) or results (e.g., Childers, 2011; Behrend, 1990). Additionally, in everyday contexts, children often see events linked to a new verb that are interspersed with distracting events. Yet most laboratory studies of verb learning show children relevant events as they hear the new verb. Thus, the present study includes distracting interleaved events to examine whether participants focus more on agents (faces) or their actions (hands) when learning verbs, and whether this varies depending on event relevancy.

To accurately acquire a verb, learners must solve what Gleitman and Gleitman (1992) described as the "packaging problem", wherein learners must package together only the relevant aspects of a dynamic event and disregard any irrelevant information. Given the difficulty of this problem, learners often use information across events and sentences or engage in cross-situational learning (e.g., Imai & Childers, 2020). Research has shown that the comparison of events can help children learn and generalize verbs (e.g., Childers & Paik, 2009; Haryu, Imai & Okada, 2011; Waxman, Lidz, Braun & Lavin, 2009). Additionally, children can benefit from seeing similar or varied events, depending on the nature of the task and test conditions (Haryu et al., 2011; Imai, Haryu, Okada, Hirsh-Pasek, Golinkoff, Shigematsu, 2008; Childers, Parrish, Olson, Burch, Fung & McIntyre, 2016).

In light of this prior research, it seems clear that young children can glean information across a set of events as they learn verbs, but theories differ in the mechanisms they posit that underlie this cross-situational processing. In statistical learning, observers compare events by noting co-occurrences between specific words and referents (e.g., Yu & Smith, 2007). In structural alignment, observers compare events linked to a new verb over instances by aligning elements across the instances based on their common relational structure (e.g., Gentner & Namy, 2006). Our study tests whether children's looking at target and distracting events during the learning phase differs in terms of their focus on agents (faces) or actions (hands). By including both types of trials, we can ask whether there are general looking preferences (e.g., for faces or hands) or whether children attend to events differently when viewing events linked to verbs vs. distracting events. To our knowledge, no prior study verb has used eye tracking in this way.

The present research is related to a set of studies with 2 ½- to 4 ½-year-old children shown the same events as in the present study (without an eye tracker) (Howard et al., 2019). Children as young as 2 ½ years were able to extend the new verb, demonstrating an ability to distinguish between target and distractor events. However, without an eye tracker, we only have indirect evidence of children's ability to ignore distracting events.

We hypothesized that children would look more to the hands as the hands AOI (area of interest) is larger (in our stimuli and in everyday life) than the face AOI, and the hands are moving. However, two prior studies demonstrate that children can also be overly attentive to the agent in an event, performing fewer verb extensions in events with multiple agents (Maguire et al., 2008), particularly when events were more complex (Childers et al., 2016). Thus, we asked whether children focused more on the face/head or the hands while seeing events in the learning phase, whether this differed depending on whether the event was linked to a new verb (target) or was not (distractor), and whether children could extend the new verbs at test.

## *1.2. Method*

### 1. Participants

Twenty-four 2 1⁄2 -year-olds (*M*age= 2;8; range: 2;0-2;11), thirty-one 3 1⁄2 -year-olds (*M*age= 3;4; range: 3;0-3;11) and twenty-one 4 ½-year-olds (*M*age= 4;4; range: 4;0-4;10) participated in this study; 40 girls, 36 boys. Children were excluded if exposure to English was <80%, if teachers reported a speech delay, for equipment failure (17) or child refusal (1).

### 2. Materials and Design

Video stimuli were created with 3 target events showing a single causative action and 2 distinct distractor actions for each of two novel verbs. Test scenes showed a new target scene and a new distractor action. A three second black screen appeared between each event.

For example, in one set, an actor picks up a natural object so that it sticks to an open hand (picking up a leaf, a stick and a rock in the target events, see Fig. 1, left), and waves a leaf around and twirls a stick on the table using her finger in the two distractor events (see Fig. 1, right). At test, children saw her pick up an object using an open hand (correct) or move an object from the center to the sides of the table (incorrect) (see Fig. 2).

Four sets of events were constructed and two of these were shown to each participant so that they could learn two verbs. Different children saw different sets to minimize the influence of a single set of events on the results. We also created three orders of the events: Target first (TTDDT), Distractor first (DTTDT) and Alternating (TDTDT), and children were assigned randomly to one of these three orders (T=Target, D=Distractor).

### 3. Experimental Set-up

Children sat in front of a 21-inch flat screen video monitor; a Tobii X30 eye tracker device was at the bottom of the monitor connected to a laptop. A webcam recorded the children's pointing responses. The distance between the table holding the monitor and participant was 16 inches. The eye tracker used a corneal reflectance tracking technique to measure eye movements. The experimenter calibrated the Tobii x30 eye-tracker using the Tobii 5-point calibration stimuli for infants; the software used was Tobii Studio.

"She's going to snarf it. She's snarfing it. She snarfed it!"

"Look what she can do!"

Fig 1: Learning phase example Target (L) and Distractor event (R).

**Test Trial 1**
"Which one is snarfing it? Can you show me? Can you point to it?

Fig 2: Test trial example.

## 4. Procedure

Two experimenters were present: one produced sentences from a script while controlling the eye tracker, the other coded children's pointing. Participants saw two warm-up trials and were asked to point to a familiar object out of a pair (e.g. "Can you point to the grapes?"). Then each child was shown two sets of events, one at a time and heard two novel verbs: gorp and snarf (one verb for each set). In each set, children saw 3 target events and 2 distractor events in one of three orders while hearing the new verbs. During the target events, they heard the novel verb three times ("She is going to ___ it. She is ___ing it. She ___ed it."). During the distractor events, they heard non-labelling speech ("Oh, look what she's doing."). At test, they saw a split screen with two different events while hearing "Now it's your turn to find <verb>ing", and were asked to point, "Point to <verb>ing. Can you point to the one who's <verb>ing?"(see Fig. 1). In a second test trial, the same videos were shown on the opposite sides of the screen "You get one more turn to play the game. Can you point to <verb>ing? Which one is <verb>ing?").

The correct side of the screen was counterbalanced. This process was repeated for a second verb. Interrater reliability calculated between the second and a third coder, both from video, showed 94% agreement with Cohen's kappa = .88 (almost perfect agreement).

### *1.3 Results*

#### 1.Pointing results

A univariate ANOVA with Age group (3: 2s, 3s, 4s) and Order (3: Target first, Distractor first, Alternating) as BS factors, dv= proportion trials correct (number correct/total trials with responses), showed a main effect of Age group, *F*(2, 75) = 3.94, *p*= .024, np2= .11, and an Age group by Order interaction, *F*(4, 75)= 3.04, p=.023, np2= .15. Given the significant interaction, we split the data by age group and repeated the univariate ANOVA within each age group, following up with one sample t-tests to compare responses to chance.

No significant effects of Order emerged in the 2 ½-year-olds and 4 ½-year-olds data. There was a significant effect of Order only in the 3 ½-year-old group, with Order, *F*(2, 30) = 4.25, *p*< .05, np2= .23. Because this result was only found in one age group, it suggests that order of the events did not exert a major effect on children's responses.

One sample t-tests showed that children in all three age groups exceeded chance. Specifically, 2 ½-year-old children's responses exceeded chance, *t*(23) = 3.40, *p*= .002, as did 3 ½-year-old children's, *t*(30) = 4.60, *p*< .001, and 4 ½-year-old children's, *t*(20) = 12.21, *p*< .001. An independent samples t-test comparing 4½–year-olds' to 3½-year-olds' responses was significant, *t*(50)= -2.25, *p*=.029 (see Fig. 2).

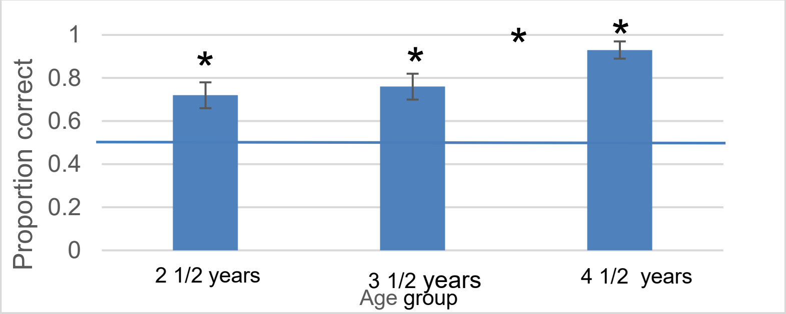


Fig. 2: Pointing results, **ps*< .05, error bars show SEM, blue line represents chance.

## 2. Eye tracking results: Looking to the face vs. the hands

A repeated measures ANOVA was computed with Age group (3: 2 ½, 3 ½, 4 ½ years) as BS factor and Trial type (2: target, distractor), Trial number (2: first, last) and AOI (2: face, hands) as WS factors; dv= total fixation duration (with zeros). There was a significant main effect of Trial type, Trial number, and AOI. There were also 3 two-way interactions, and a 3 way interaction of Trial type x Trial number x AOI, *F*(1, 67)=27.36, *p*< .001.

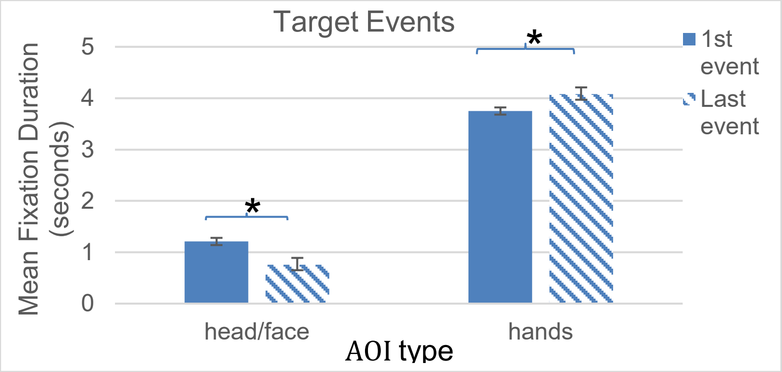


Fig. 3. Target Events: Graph shows mean total fixation duration by Trial (first, last) and AOI type (face, hands), **ps*< .05.

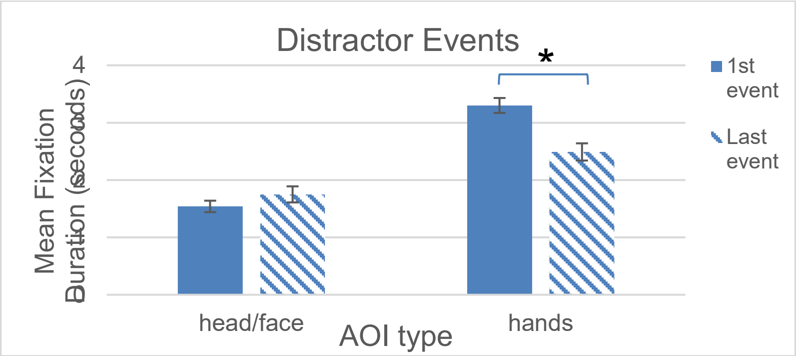


Fig. 4. Distractor Events: Graph shows mean total fixation duration by Trial (first, last) and AOI type (face, hands), **p*< .05.

<u>Target Trials:</u> A repeated measures ANOVA examining looking during the Target trials revealed a main effect of AOI, *F*(1, 69)= 580.62, *p*< .001, and a Trial number x AOI interaction, *F*(1, 69)= 30.14, *p*< .001. Pairwise

comparisons with Sidak corrections showed that, in both the First and Last Target events, children looked longer at the Hands than the Face, *p*s< .05. Importantly though, across trials, looking to the Hands increased, *p*= .003, while looking to the Face significantly decreased, *p*< .001 (see Fig. 3).

<u>Distractor trials:</u> A repeated measures ANOVA revealed a main effect of Trial number, *F*(1, 69)= 26.96, *p*< .001, AOI, *F*(1, 69)= 53.91, *p*< .001, and a Trial number x AOI interaction, *F*(1, 69)= 14.10, *p*< .001. Pairwise comparisons with Sidak corrections showed that children looked longer at the Hands than the Face in the first distractor event, *p*< .001, and in the second distractor event, *p*=.004. However, across trials, a different pattern was seen with children's looking to the Hands decreasing, *p*< .001, while looking to the Face was maintained (see Fig. 4).

### *1.4 Discussion*

Children learning verbs often see other intervening events as they are learning verbs (e.g., seeing a stirring event while learning the verb 'chop' in the kitchen), which need to be processed differently than events linked to the target verb. Our results provide experimental evidence for differential visual processing of relevant vs. distractor events as events are experienced. Specifically, across age groups, when viewing relevant events, children increase their looking to the hand region (actions) over trials and decrease their looking to the agents' face, which is less informative for learning a verb's meaning. In contrast, when viewing distracting events, children decrease their looking to hands over trials and maintain their attention to the head. These results add to the body of research showing that children can compare events during verb learning, and that comparisons help them extend new verbs (e.g., Childers et al., 2016; Scott & Fisher, 2012; Waxman et al., 2009). Specifically, they provide evidence that children are adjusting their visual fixations differently as they see relevant events linked to a new verb as opposed to irrelevant events, suggesting that they are strategic in how they visually inspect events.

Learning new verbs is important to learning one's native language. Our study suggests that by 2 ½ years, children have developed visual strategies for inspecting events that should help them attend to events appropriately when seeing relevant events and hearing verbs (attending more to what the hands are doing than the face) and perhaps help them ignore distracting information (as they focus on faces and not hands) when they see irrelevant events. These are exciting new findings that reveal what mental mechanisms could underlie early verb learning. The study of children's acquisition of verbs, which vary across languages, could suggest ways languages may have changed in response to the learnability of emerging predicate structures.

## Acknowledgements

Funding for this research was provided by the National Institute of Health (2R15 HD044447), Murchison Scholars Fund, and Trinity University. We thank Victoria Bell, Sophia Arriazola, Aidan Burke, Priscilla Tovar-Perez, Hannah Bortz, Katy Capps, and Sophia Rodriguez for data collection and coding.

## References

Behrend, D. A. (1990). The development of verb concepts: Children's use of verbs to label familiar and novel events. *Child Development*, *61*, 681-696.

Childers, J. B. (2011). Attention to multiple events helps two-and-a-half-year-olds extend new verbs. *First Language, 31(1),* 3-22.

Childers, J. B., & Paik, J. H. (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language, 36(1)*, 201–224.

Childers, J. B., Parrish, R., Olson, C. V., Burch, C., Fung, G., & McIntyre, K. P. (2016). Early verb learning: How do children learn how to compare events? *Journal of Cognition and Development, 17(1)*, 41–66.

Childers, J. B., Paik, J. H., Flores, M., Lai, G., & Dolan, M. (2017). Does variability across events affect verb learning in English, Mandarin, and Korean? *Cognitive Science, 41,* 808-830.

Gentner, D., & Namy, L. L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science, 15(6)*, 297–301.

Gleitman, L. R., & Gleitman, H. (1992). A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition. *Current Directions in Psychological Science, 1(1),* 31-35.

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action based verb extension. *Child Development, 82(2),* 674-686.

Howard, T. J., Porter, B. M., Childers, J. B. (2019). Can young children ignore irrelevant events, or subevents, during verb learning? Journal of Cognition and Development, 20(3), 411-432.

Imai, M., & Childers, J. B. (2020). Learning individual verbs and the verb system: When are multiple examples helpful? In J. B. Childers (Ed.), *Language and Concept Acquisition from Infancy Through Childhood* (pp. 131-158). Springer, Cham.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development, 79(4),* 979-1000.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science, 11(4),* 628–634.

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition, 122(2),* 163–180.

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology, 59(1),* 67–95.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18(5),* 414–420.

# EVIDENCE FOR THE COGNITIVE ORIGINS OF SYNTACTIC HARMONY

JENNIFER CULBERTSON*[1] and SIMON KIRBY[1]

*Corresponding Author: jennifer.culbertson@ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, United Kingdom

## 1. Background

There is a consensus that cognition-general mechanisms and biases shape syntactic typology. However there is very little direct behavioral evidence linking such mechanisms to specific typological patterns in syntax. Here we test a recent proposal by Culbertson and Kirby (2016) that syntactic harmony–alignment between syntactic heads and dependents (Greenberg, 1963)–reflects a domain-general bias for simplicity (Chater & Vitányi, 2003), acting on linearized, language-specific categories (i.e., heads and dependents). Previous research suggests that human learners do indeed prefer syntactic harmony (e.g., Culbertson, Franck, Braquet, Barrera Navarro, & Arnon, 2020). However, if harmony is the result of a domain-general bias for simplicity, then a similar preference should hold when sequences of *meaningless and/or non-linguistic* categories are linearized.

## 2. Experiment 1: meaningless letter sequences

In Experiment 1 we take one step away from syntactic harmony by using sequences of meaningless letter strings rather than combinations of meaningful words. Stimuli consisted of two categories: ‘head’ elements and ‘dependent’ elements. In natural language these categories are presumably learned based on similarities in structure and/or meaning. Here, the categories are distinguished based on length and structure. Heads are comprised of contrasting CVCVC strings, and dependents of CVC strings. Dependents paired with each type of head have C’s that match the heads in terms of voicing, and V’s chosen from the same set in order to help learner identify the relevant head-dependent combinations (akin to different phrase types in natural language).

Participants (N=74) were randomly assigned to one of three conditions: harmonic (either Head-Dep or Dep-Head), non-harmonic across heads (Head-Dep order for one head type, Dep-Head for the other), and non-harmonic within heads (one Dep type for each Head type was Dep-Head, the other Head-Dep). They were exposed to grammatical sequences in the language, and then tested using 2AFC. Accuracy was highest in the harmonic condition (vs. non-harmonic across heads $\beta = -3.18 \pm 0.81, p < 0.001$; vs. non-harmonic within heads $\beta = -5.70 \pm 0.83, p < 0.001$) and lowest in the non-harmonic within heads condition (vs. non-harmonic across heads $\beta = -2.38 \pm 0.55, p < 0.001$).

## 3. Experiment 2: paired abstract shapes

Experiment 2 was identical to Experiment 1 in terms of the structure of the "languages" and the experimental conditions participants (N=76) were assigned to. The only difference was in how head and dependent elements were constructed: heads were single larger shapes, dependents were two smaller stacked shapes, positioned either to the left or right of the larger shape. As in Experiment 1, dependents paired with each head matched, here in the roundness of the shapes in order to help learners identify the relevant head-dependent combinations. Accuracy was highest in the harmonic condition (vs. non-harmonic across heads $\beta = -1.88 \pm 0.72, p < 0.001$; vs. non-harmonic within heads $\beta = -4.02 \pm 0.71, p < 0.001$) and lowest in the non-harmonic within heads condition (vs. non-harmonic across heads $\beta = -1.92 \pm 0.51, p < 0.001$).

Figure 1. Head and dependent stimuli.

Figure 2. Experiment 1.

| Heads | Dependents |
|---|---|
| H1: nageng, negang, | Dep1a: bav, baz, dav, daz |
| genang, ganeng | Dep1b: veb, ved, zeb, zed |
| H2: shukoth,shokuth, | Dep2a: puf, pus, tuf, tus |
| koshuth, kushoth | Dep2b: fop, fot, sop, sot |

Figure 3. Experiment 2.

| Heads | Dependents |
|---|---|
| H1:  | Dep1a:  |
| | Dep1b:  |
| H2:  | Dep2a:  |
| | Dep2b:  |

## 4. Discussion

In language, we see a preference for harmony in the linearization of grammatical categories—heads and dependents comprising syntactic phrases. Here we have shown that a linearization preference akin to harmony will emerge in similarly structured stimuli when no linguistic meaning is present, and even when the stimuli are non-linguistic in nature. Our findings suggest that rather than being language-specific, the origins of harmony lie in a cognition-general bias for simpler representations, active in individual learners, likely amplified over time via cultural transmission.

## References

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.

Culbertson, J., Franck, J., Braquet, G., Barrera Navarro, M., & Arnon, I. (2020). A learning bias for word order harmony: evidence from speakers of non-harmonic languages. *Cognition*, *204*.

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain general biases have domain specific effects. *Frontiers in Psychology*, *6*(1964).

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (p. 73-113). Cambridge, MA: MIT Press.

# NON-ARBITRARINESS ACROSS MODALITIES: BOOTSTRAPPING THE EMERGENCE AND DEVELOPMENT OF QUESTIONS

CHRISTINE CUSKLEY[1], JOHANNES M. HEIM[*1,2]

*Corresponding Author: johannes.heim@abdn.ac.uk
[1]Language Evolution, Acquisition, and Development Group, Newcastle University, Newcastle upon Tyne, UK
[2]Language and Linguistics, University of Aberdeen, Aberdeen, UK

Over the last decade, language evolution research has been the driving force behind a renewed interest in non-arbitrary mappings (Cuskley & Kirby, 2013), as well as multimodal theories of language development (Vigliocco et al., 2014). These theories have focused largely on lexical iconicity in language systems. Here, we begin to extend this line of research by considering how non-arbitrariness plays a role in language development and emergence beyond the lexicon, moving from lexical to pragmatic meaning across different modalities. In other words, we focus on non-arbitrariness in situated language (Murgiano et al., 2021; Özyürek, 2021). Specifically, we consider the role of indices of dominance in scaffolding the development of questions. By surveying the state of the art in research on prosody, gesture, and signing, we outline evidence for cross-linguistically attested, iconic gestural scaffolding of questions across signed and spoken languages. This supports development of indexical use of pitch in question prosody, which has analogies in non-arbitrary dominance signalling. We conclude by synthesizing this evidence, presenting a novel non-arbitrary bootstrapping theory for the emergence of questions.

## 1. Introduction

Making our desires known through linguistic directives, such as questions, lies at the very heart of human communication. Questions play a key role in driving conversations by engaging in a collaborative effort in building common ground (Roberts, 2012) and have a prominent place in accounts of first language acquisition (Thornton, 2016). While morphosyntactic cues to questions are arbitrarily aligned and exhibit considerable cross-linguistic variation, different modalities – gesture and intonation - are aligned across many languages. For example, while both English and Mandarin use rising intonation to encode questions, English combines rising intonation with a sentence-initial auxiliary ‘do’ and Mandarin with a question particle. Moreover, questions in both languages will be accompanied by raised eyebrows, hands, and shoulders.

Intonation and by extension, gesture (Bolinger, 1983) can mirror features associated with dominance asymmetries (Ohala, 1983). We argue that questions inherently involve a temporary dominance asymmetry within a conversational dyad, where the speaker uses upward movement to encode that they need something from the addressee (materially or informationally).

We propose that a rich inventory of developmentally robust and cross-culturally attested multimodal, that is: non-morphosyntactic cues scaffolds the emergence of complex linguistic questions, both in development and language evolution. This paper will draw on previous literature for synthesising three distinct lines of evidence for this novel theory: i) the cross-linguistic ubiquity of non-morphosyntactic cues to questionhood ii) the importance of such cues in early pragmatic development, iii) the parallels of grammaticalized facial cues in spoken and signed languages. On this basis, we show that non-arbitrary cues are likely to have played a key role in the emergence of questions given them a role as a catalyst in language development, the key features occurring consistently across languages and dominant signalling modalities.

## 2. Non-arbitrariness across modalities and languages

Questions of whatever morphosyntactic form engage the addressee to respond to the speaker's epistemic stance (Heim, 2019). Questions are widely considered the marked form of human conversation as opposed to assertions, which occur more frequently (Huddleston, 1994**.** Correspondingly, an auxiliary-initial question word order, which make up the majority of questions in the caregiver input of an English-speaking child (Estigarribia 2009), occurs less often than an auxiliary-medial word order (Newport et al., 1977). Morphosyntactic marking of questions is present in 80% of the world's languages (Dryer, 2008), and comes in the shape of word order changes, interrogative clitics and particles, and/or verbal inflection (Siemund, 2001)**.** Content questions have an additional cue form of a variable corresponding to the element unknown (e.g., *wh*-pronouns in English) and sometimes omit other cues present in yes-no questions. While morphosyntactic marking varies significantly across languages, other modalities, such as intonation and gesture, show considerable cross-linguistic consistency.

By far the most common cue to mark questions is rising intonation (Ultan, 1978), although prosodic details vary alongside how pitch is employed at the lexical level. Pitch plays a key role in other kinds of non-arbitrariness, particularly linked to size sound symbolism via the frequency code hypothesis (Ohala, 1994). The mechanism of motivatedness behind the frequency code is that a larger resonant chamber is an index for a larger producer, but motivatedness can have multiple mechanisms (Ahlner & Zlatev, 2010). For example, while high vowels are generally considered "smaller" than low vowels, Diffloth (1994) describes Bahnar, a language where high vowels are large and low vowels are small, mediated by the size of the tongue within the mouth (which occupies a larger volume when producing high vowels like /i/ relative to low vowels like /a/).

Despite some cross-linguistic variation in how question intonation is deployed, it is nonetheless a reliable indexical cue, motivated by dominance asymmetries in pragmatic context across languages. For example, Gussenhoven & Chen (2000) show that peak alignment, peak height, and end pitch are associated with questions by speakers of Dutch, Chinese, and Hungarian; although these languages vary in their specific encoding of questions. In general, animals and humans exhibit a strong association between pitch and dominance (Puts et al., 2007), with lower pitch signalling larger size and greater dominance.

The manual-gestural modality shows similar non-arbitrariness in marking questions, arguably tied more straightforwardly to dominance: hands and eyebrows are raised (Krahmer & Swerts, 2005). This contrasts markedly with negation, where facial articulators show the opposite direction: eyebrows are lowered, and lips are pressed against each other (Benitez-Quiroz et al., 2016). Hence, just like with intonation, a downward movement expresses dominance, an upward movement a lack thereof. One other prevalent gesture associated with the absence of knowledge is the palm-up (epistemic) gesture, which is also associated with submissiveness (Cooperrider et al., 2018) and echoes the upward pitch contours often associated with questions in speech (Ultan, 1978). In contrast to co-speech manual gestures, House (2002) argues that facial cues are less reliable than manual ones for encoding questions, at least in spoken languages. Sign languages, however, are more reliable in this context: They generally mark questions with raised eyebrows (Cecchetto, 2012) including ASL, Sign Language of the Netherlands, and Kazakh-Russian Sign Language (Kimmelman et al., 2020). This facial gesture is also a widespread marker of submissiveness in primates (Keating, 1985), lending further support to the idea that dominance indices played a catalyst role in language emergence.

Areas that require further research are cues such as breathiness, duration effects**,** and speech rate (Dehé et al., 2022)**,** which have all been found to encode questions in several languages. Moreover, in Yélî Dnye (Levinson, 2010), questions are marked by mutual gaze making rather than intonation. This nonetheless arises from systematic dominance asymmetries within a conversation (with the speaker asserting dominance, rather than telegraphing submission).

## 3. Non-arbitrariness in pragmatic development

Non-arbitrary cues to questionhood are not only attested widely across languages and modalities, but also in language development. Here, too, non-arbitrary cues tend to be non-morphosyntactic. Intonation and gesture provide an avenue into pragmatic development, and questions in child directed speech which use non-morphosyntactic cues are ubiquitous (Kania, 2016). Morphosyntactic encoding of questions, on the other hand, has a long developmental trajectory, with the first questions marked by word order occurring from around 20 months in Italian, German and Swedish, but only from 30 months in English. It takes years for children to master the morphosyntactic variation in questions, especially when they include negation (Guasti, 2017). With non-morphosyntactic cues appearing early, the acquisition of morphosyntactic cues can benefit from multimodal scaffolding in development, building on motivated manual and facial gestures, which then bootstrap indexical use of pitch.

We assume manual and facial gestures to be the starting point of this developmental process, because they are the first means available for communication (Iverson & Goldin-Meadow 2005). Gestures frequently serve as precursors for what is later expressed using more complex linguistic structure (Capirci et al., 2005). Children then combine vocalizations and gestures for relating non-arbitrary signs to previous experience with a rapid increase of gestures during the first year (Burkhardt-Reed et al., 2021). Pointing, too, serves as a cue to questions when the addressee is knowledgeable (Begus & Southgate, 2012). From 14 months, infants use prelinguistic vocalizations to differentiate conversational strategies (Grünloh & Liszkowski, 2015). Combining these findings with the those of intonational cues, we find that non-arbitrary forms are well-suited for bootstrapping linguistic development (Bohn et al., 2019).

Intonational cues follow early on. Sensitivity to different sentence melodies is evident form as early as 5 months (Frota et al., 2014). Intonational distinctions of questions from other speech acts occur as early as 7 months, and adult-like intonation is produced by English-speaking children in their second year of life (Prieto Vives & Esteve Gibert, 2013). Yet, infants show notable variation as to when this command equals adult performance. Correspondingly, children use different strategies for asking questions up to age 11 (Patel & Grigos, 2006). German children aged 2;8-2;10 (age; month) show striking variability in producing question intonation and appear not to reach prosodic targets before age 3;0 (Lleó & Rakow, 2011). Hübscher and colleagues (2019) show that facial and prosodic cues for expressing knowledge states are employed at the age of three, while lexical items are only employed later. This strongly suggests that non-arbitrary, non-morphosyntactic cues play a central role in advancing pragmatic development and can act as a platform for later integrating morphosyntactic cues.

## 4. Non-arbitrariness in signed and spoken languages

Parallels in the non-arbitrariness of encoding questions across spoken and signed languages add a third piece of evidence supporting our hypothesis of the catalytic function of these cues. Broadly speaking, non-arbitrariness has been shown to play an important role across both infant-directed speech and signing (Brand et al., 2002) across different modalities. Perniss and colleagues (2015) demonstrate that non-arbitrariness plays a greater role in referring to non-visible referents in child-directed signing compared to adult-directed speech. Laing and colleagues (2017) show that prosody in onomatopoetic words (e.g., *quack*) is more marked in infant-directed speech than in equivalent conventional words (e.g., *duck*). In both studies, however, the focus was very much on the non-arbitrary signalling of lexical meaning, where the benefits for language development are well-established (Ota et al., 2018; although see Nielsen & Dingemanse, 2021). This suggests that children acquiring language – regardless of the dominant modality (speech or sign) – exploit non-arbitrariness. We argue this extends to cues which aid in acquiring pragmatic concepts, such as questions.

As in section 2, we focus here on the directionality of these cues, which shows consistent trends across spoken and signed languages. Zeshan (2004) reports that most signed languages use a combination of eyebrow raise, eyes wide open, eye contact, a head forward position or forward body posture to ask questions. These non-manual interrogative signs accompany manual/lexical question markers, such as content question words or dedicated morphemes. The non-manual signs are key components of marking questions across many sign languages (Kimmelman et al., 2020). Their frequent presence in questions remind us that these are the marked speech acts compared to assertions. Remarkably, the directionality of these upward and forward oriented non-manual signs in signed languages mirrors that of the intonational patterns observed in many spoken languages (Dixon, 2012). The same applies for the upward movement of co-speech gestures in spoken languages.

Even more remarkable is the fact that directionality changes exhibit the same patterns across modalities. Content and polarity questions in English, for instance, have a fall and rise as their standard intonational contours, respectively (Bartels, 1997). Reversing the directionality for these types is meaningful, too, as their function changes from and open to a closed question (Hedberg et al., 2017), where 'closedness' evokes a set expectation about the answer. We find the analogous pattern for non-manual signs in sign-languages. Italian Sign Language has raised eyebrows in polar questions and lowered eyebrows in content questions (Zeshan, 2004) just as spoken Italian has a falling-rising contour for polar questions and a falling contour for content questions (D'Imperio, 2002). The paradigm is complete with lowered eyebrows reported for closed or loaded questions, as reported for Israeli Sign Language (Meir, 2004). The similarity between spoken and signed questions also has some equivalence in morphosyntax where both signed and spoken languages use the words for indefinite pronouns and question markers (Dixon, 2012). Studies that address the role of non-arbitrariness across infant-directed signing and speaking with respect to interactional strategies are still in their infancy and researchers have just begun the explore the richness of multimodal cues in the learner's input. (Murgiano et al., 2021)

## 5. Conclusions

By comparing non-morphosyntactic cues across languages, ages, and modalities, we have argued that their non-arbitrariness presents an opportunity for bootstrapping morphosyntactic development of questions. By drawing on a wide range of findings from production and perception data, we establish that non-morphosyntactic cues are likely to be non-arbitrary, transcend cultural and linguistic borders, and precede morphosyntactic competence in infancy. These properties make non-arbitrary cues to questionhood a strong candidate for playing a central role in pragmatic development, and by extension in language emergence due to their prevalence and early availability. Moreover, this theory opens specific hypothesis spaces which can be probed with more specific experimental and typological work. For example, using iterated artificial language learning, we might expect that readily available indexical cues might outcompete complex morphosyntactic cues for questions in a dyadic face to face context. We outline a possible experimental approach wherein participants are trained on an artificial language with complex morphosyntactic cues for questionhood used irregularly (i.e., in 60% of questions) in the first generation. In a face-to-face condition, we predict that these cues will either stagnate or fade out of the system over iterated "generations", in favor of indexical cues. In a virtual condition without the availability of indexical cues, we predict the cues will regularize and come to be more reliable cues to questionhood. In addition, typologically, we might expect that the 20% of languages without complex morphosyntactic cues (Dryer, 2008) are likely to have richer non-morphosyntactic cues. Assuming non-arbitrary, non-morphosyntactic cues are to some extent present even in the many languages with complex morphosyntactic marking, this points to the former type of cues as an evolutionarily older strategy. In summary, we have established (i) the consistent cross-linguistic availability of non-arbitrary question cues, (ii) the use of these cues in bootstrapping the acquisition of question, and (iii) evidence that similar non-arbitrary cues play an important role in both acquisition and evolution. Together, these lines of evidence point to a novel theory for the emergence of questions, where non-arbitrary cues from intonation, gesture, and signing play a catalyst role in the language emergence and development.

## References

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*(1/4), 298–348.

Bartels, C. (1997). *Towards a compositional interpretation of English statement and question intonation*. University of Massachusetts Amherst.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617.

Benitez-Quiroz, C. F., Wilbur, R. B., & Martinez, A. M. (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, *150*, 77–84.

Bohn, M., Call, J., & Tomasello, M. (2019). Natural reference: A phylo- and ontogenetic perspective on the comprehension of iconic gestures and vocalizations. *Developmental Science*, *22*(2), e12757.

Bolinger, D. (1983). Intonation and gesture. *American Speech*, *58*(2), 156–174.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1),

Burkhardt-Reed, M. M., Long, H. L., Bowman, D. D., Bene, E. R., & Oller, D. K. (2021). The origin of language and relative roles of voice and gesture in early communication development. *Infant Behavior and Development*, *65*, 101648.

Capirci, O., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture*, *5*(1–2), 155–177.

Cecchetto, C. (2012). Sentence Types. In *Sign Language: An international handbook* (pp. 292–315). De Gruyter Mouton.

Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign. *Frontiers in Communication*, *3*.

Cuskley, C., & Kirby, S. (2013). Synesthesia, cross-modality, and language evolution. *Oxford Handbook of Synaesthesia*, *20*, 869–907.

Dehé, N., Braun, B., Einfeldt, M., Wochner, D., & Zahner, K. (2022). The prosody of rhetorical questions: A cross-linguistic view. *Linguistische Berichte*, *269*, 3–42.

Diffloth, G. (1994). I: Big, a: Small. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 107–130). CUP.

D'Imperio, M. (2002). Italian intonation: An overview and questions. *Probus*, *14*(1).

Dixon, R. M. (2012). *Basic linguistic theory, vol. 3: Further grammatical topics*. Oxford: OUP.

Dryer, M. S. (2008). Polar Questions. In *World Atlas of Language Structures*. Max Planck Digital Library. http://wals.info/feature/116

Frota, S., Butler, J., & Vigário, M. (2014). Infants' Perception of Intonation: Is It a Statement or a Question? *Infancy*, *19*(2), 194–213.

Grünloh, T., & Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts*. *Journal of Child Language*, *42*(6), 1312–1336.

Guasti, M. T. (2017). *Language acquisition: The growth of grammar*. MIT press.

Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. *6th International Conference on Spoken Language Processing (ICSLP 2000)*, 91–94.

Hedberg, N., Sosa, J. M., & Görgülü, E. (2017). The meaning of intonation in yes-no questions in American English: A corpus study. *Corpus Linguistics and Linguistic Theory*, *13*(2), 321–368.

Heim, J. M. (2019). *Commitment and engagement: The role of intonation in deriving speech acts* [PhD Thesis]. University of British Columbia.

House, D. (2002). Perception of question intonation and facial gestures. *TMH-QPSR Fonetik*, *44*, 41–44.

Hübscher, I., Garufi, M., & Prieto, P. (2019). The development of polite stance in preschoolers: How prosody, gesture, and body cues pave the way. *Journal of Child Language*, *46*(5), 825–862.

Huddleston, R. (1994). The contrast between interrogatives and questions. *Journal of Linguistics*, *30*(2), 411–439.

Kania, U. (2016). *The Acquisition and Use of Yes-no Questions in English: A Corpus-Study from a usage-based perspective* (Vol. 36). Narr Francke Attempto.

Keating, C. F. (1985). Human Dominance Signals: The Primate in Us. In S. L. Ellyson & J. F. Dovidio (Eds.), *Power, Dominance, and Nonverbal Behavior* (pp. 89–108). Springer.

Kimmelman, V., Imashev, A., Mukushev, M., & Sandygulova, A. (2020). Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study. *PLOS ONE*, *15*(6), e0233731.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*(Pt 1), 29–53.

Laing, C. E., Vihman, M., & Keren-Portnoy, T. (2017). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *Journal of Child Language*, *44*(5), 1117–1139.

Levinson, S. C. (2010). Questions and responses in Yélî Dnye, the Papuan language of Rossel Island. *Journal of Pragmatics*, *42*(10), 2741–2755.

Lleó, C., & Rakow, M. (2011). Intonation targets of yes/no questions by Spanish and German monolingual and bilingual children. In E. Rinke & T. Kupisch (Eds.), *Hamburg Studies on Multilingualism* (Vol. 11, pp. 263–286). John Benjamins.

Meir, I. (2004). Question and Negation in Israeli Sign Language. *Sign Language & Linguistics*, *7*(2), 97–124.

Murgiano, M., Motamedi, Y., & Vigliocco, G. (2021). Situating Language in the Real-World: The Role of Multimodal Iconicity and Indexicality. *Journal of Cognition*, *4*(1), 38.

Newport, E. L., Gleitman, H. & Gleitman, L. R. (1977). Mother I'd rather do it myself: some effects and non-effects of maternal speech style. In Snow, C. E. & Ferguson, C. A... (eds), Talking to children: language input and acquisition. Cambridge: C.U.P.

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in Word Learning and Beyond: A Critical Review. *Language and Speech*, *64*(1), 52–72.

Ohala, J. (1983). The Origin of Sound Patterns in Vocal Tract Constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). Springer.

Ohala, J. (1994). The frequency code underlies the symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound Symbolism* (pp. 325–347). CUP.

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why Choo-Choo Is Better Than Train: The Role of Register-Specific Words in Early Vocabulary Growth. *Cognitive Science*, *42*(6), 1974–1999.

Özyürek, A. (2021). Considering the Nature of Multimodal Language from a Crosslinguistic Perspective. *Journal of Cognition*, *4*(1), 42.

Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question–statement contrast in 4, 7 and 11 year-old children. *Speech Communication*, *48*(10), 1308–

Perniss, P., Zwitserlood, I., & Özyürek, A. (2015). Does space structure spatial language? *Language*, *91*(3), 611–641.

Prieto Vives, P., & Esteve Gibert, N. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research. 2013 Jun; 56: 850-64.*

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution & Human Behavior*, *28*, 340–344.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*, 6–1.

Siemund, P. (2001). Interrogative constructions. *Language Typology and Language Universals*, *2*, 1010–1028.

Thornton, R. (2016). *Acquisition of Questions* (J. L. Lidz, W. Snyder, & J. Pater, Eds.; Vol. 1). Oxford University Press.

Ultan, R. (1978). Some general characteristics of interrogative systems. In J. Greenberg (Ed.), *Universals of Human Language* (pp. 83–124). Stanford UP.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130292.

Zeshan, U. (2004). Interrogative Constructions in Signed Languages: Crosslinguistic Perspectives. *Language*, *80*(1), 7–39.

# LEARNABILITY AND EMERGENCE OF DEPENDENCY STRUCTURES IN AN ARTIFICIAL LANGUAGE

Emily Davis[*1] and Kenny Smith[2]

[*]Corresponding Author: e4davis@ucsd.edu
[1]Department of Linguistics, UC San Diego, La Jolla, CA, USA
[2]Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, UK

In a pair of artificial language experiments, we investigated the learnability and emergence of different dependency structures: branching, center-embedding, and crossed. In natural languages, branching is the most common dependency structure; center-embedding occurs but is often disfavored, and crossed dependencies are very rare. Experiment 1 addressed learnability, testing comprehension and production on small artificial languages exemplifying each dependency type in noun phrases. As expected, branching dependency grammars were the easiest to learn, but crossed grammars were no harder to learn than center-embedding. Experiment 2 employed iterated learning to examine the emergence and stabilization of consistent grammar using the same type of stimuli as Experiment 1. The initial participant in each chain of transmission was trained on phrases generated by a random grammar, with the language produced by that participant passed to the next participant through an iterated learning process. Branching dependency grammar appeared in most chains within a few generations and remained stable once it appeared, although one chain stabilized on output consistent with a crossed grammar; no chains converged on center-embedding grammars. These findings, along with some previous results, call into question the assumption that crossed dependencies are more cognitively complex than center-embedding, while confirming the role of learnability in the typology of dependency structures.

## 1. Introduction

In a pair of artificial language experiments, we investigated the learnability of different dependency structures: branching, center-embedding, and crossed. Long-distance dependencies between words (e.g. nouns and verbs, or nouns and adpositions) are an essential aspect of human language, and can be arranged in different ways. In (1), the dependency between subject and verb is indicated by the subscript (examples based on Vosse and Kempen 1991):

1.

a. (Right-) branching dependencies in English:
… when John$_1$ saw$_1$ Peter$_2$ walk$_2$

b. Center-embedded dependencies in German:
… als Johan$_1$ Peter$_2$ laufen$_2$ sah$_1$

c. Crossed dependencies in Dutch:
… toen Jan$_1$ Peter$_2$ sag$_1$ lopen$_2$

Most natural languages predominantly use some combination of branching and center-embedding dependency structures; crossed dependencies are rare, but some examples are attested, e.g. in Dutch subordinate clauses (Bresnan et al. 1982). Center-embedding tends to be less prevalent than branching both within and across languages, likely because of the cognitive difficulties it presents (see e.g. Kuno 1974, Hawkins 2004). Some experimental evidence from natural language processing (Bach et al. 1986) and nonlinguistic sequence learning (Öttl et al. 2015) also suggests that crossed dependencies are easier to process and learn than center-embedded dependencies, which is surprising in light of the marginal status of crossed dependencies in natural languages.

Artificial language learning experiments can shed light on the cognitive factors shaping natural language structures (e.g. Culbertson 2012, Fedzechkina 2018), and the regularization and stabilization of irregular input (e.g. Hudson Kam & Newport 2005). In the present study we applied these methods to dependencies. In two experiments, we investigated the learnability of different dependency types within the same paradigm, using meaningful sequences (descriptions for visual scenes) that constituted a small artificial language with several nouns and spatial adpositions.

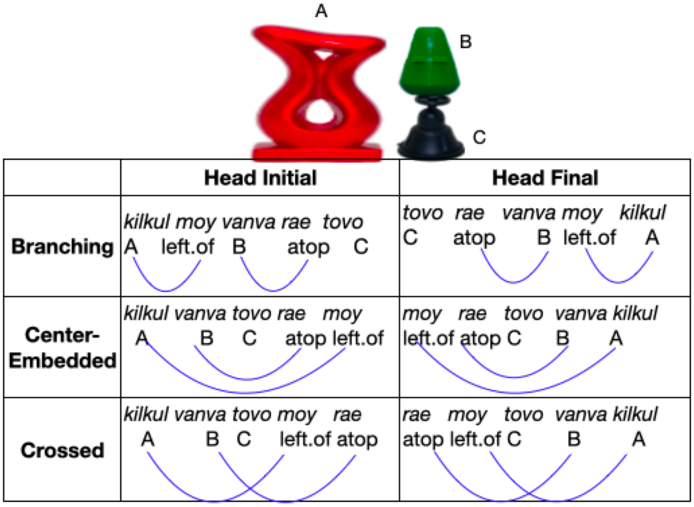
## 2. Experiment 1: Learnability

Experiment 1 concerned the learnability of the three types of dependency structures, as well as head-dependent order (head-initial or head-final), using noun phrases with locative adpositional phrases for configurations of novel objects.

### *2.1. Methodology*

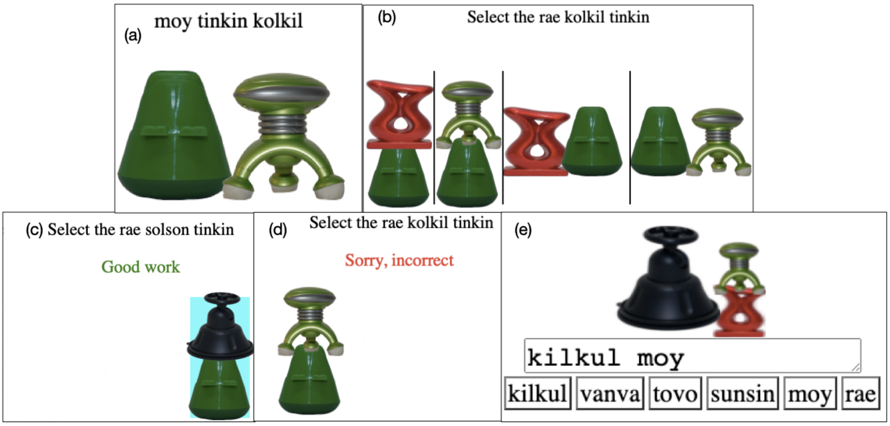
We created simple artificial languages representing the six possible combinations of dependency type and order (Figure 1). The associated visual stimuli were composed of images from the NOUN database as components (Horst 2015). Both experiments were conducted online through Mechanical Turk. Participants (n = 120, 20 per condition, all self-reported English speakers) were trained on one of these languages by a mix of passive training trials (viewing individual objects or configurations of objects with an appropriate description; Figure 2a) and comprehension trials (being presented with a description and attempting to identify the corresponding scene from an array of possibilities, with corrective feedback provided; Figure 2b-d).

Over the course of the training procedure participants progressed from scenes involving single objects to three-object configurations. In the final stage of the experiment, participants were asked to produce descriptions of scenes by clicking on labeled buttons (Figure 2e); these final test items included new configurations not seen in training (including some scenes featuring four objects), in order to assess generalization.



|  | Head Initial | Head Final |
|---|---|---|
| **Branching** | *kilkul moy vanva rae tovo*<br>A left.of B atop C | *tovo rae vanva moy kilkul*<br>C atop B left.of A |
| **Center-Embedded** | *kilkul vanva tovo rae moy*<br>A B C atop left.of | *moy rae tovo vanva kilkul*<br>left.of atop C B A |
| **Crossed** | *kilkul vanva tovo moy rae*<br>A B C left.of atop | *rae moy tovo vanva kilkul*<br>atop left.of C B A |

Figure 1. Sample array with description in each of the six syntactic orders in Experiment 1. Objects are labeled with letters for clarity here and in other example arrays; no such labels were used on stimuli for the actual experiment.



moy tinkin kolkil

Select the rae kolkil tinkin

Select the rae solson tinkin

Good work

Select the rae kolkil tinkin

Sorry, incorrect

kilkul moy

Figure 2: Passive training (a), comprehension (b), feedback on correct (c) and incorrect responses to comprehension (d) trials, and production trial with partially entered description (e) from Experiment 1.

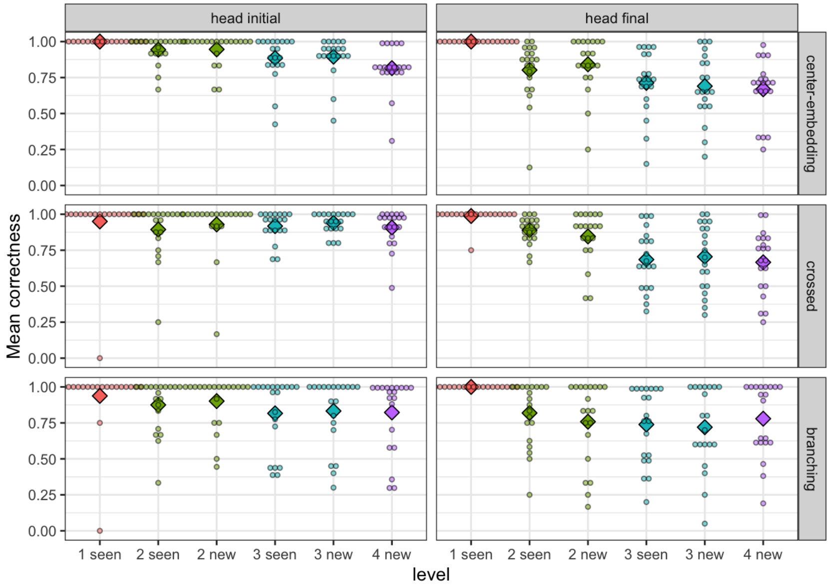


Figure 3: Learning accuracy (1 - string edit distance between trained and produced descriptions) by level (number of objects in the scene) and grammar type. Each point is one participant's performance at that level; large diamonds represent mean performance across all participants. "Seen" indicates items seen in training; "new" are novel test items. Note that participants were not tested on single-item scenes, and were not trained on 4-item scenes.

## ***2.2. Analysis and results***

Participant-created labels were analyzed for consistency with the training data (evaluated by calculating string edit distance between the trained and produced label for each scene) and for adherence to one of many possible grammars (by generating grammars with all possible combinations of dependency order and headedness, and finding the grammar which generated the description which best matched the participant's description, as assessed by edit distance). Figure 3 shows accuracy by condition. As expected, branching grammars were the easiest to learn: learning accuracy on the mean of center-embedded and crossed dependencies was significantly lower than that for branching structures as shown by a linear mixed effects model (B = -0.02, SE = 0.001, p = .04), and participants trained on branching grammars generally produced output consistent with a branching grammar. Learning accuracy for crossed grammars was not significantly different from center-embedding grammars (B = -0.01, SE = 0.02, p = .51). However, 6 participants of the 40 trained on center-embedded grammar produced labels consistent with a crossed grammar, suggesting that they had reanalyzed the input in conformity with a new grammar. The converse (participants trained on crossed grammar generating center-embedded strings) did not occur. There was also a significantly greater decline in accuracy with level in head-final than in head-initial conditions (B = -0.03, SE = 0.01, p < .001).

## 3. Experiment 2: Emergence of grammar

Experiment 2 set out to examine the emergence and stabilization of a consistent grammar in noun phrases, and furthermore to observe whether the results reflected real language typology. We hypothesized that branching syntax would be predominant in the results, due to its greater learnability and comprehensibility, whereas center-embedding would be uncommon and crossed grammar might appear rarely, in keeping with previous experimental results (e.g. Öttl et al. 2015) and with the learnability results of Experiment 1. To this end we employed the method of *iterated learning* (Kirby et al. 2008), using stimuli of the same type as Experiment 1. Iterated learning is a process where the response of one participant (in our case, the descriptions produced for scenes) becomes the training input for the next participant in a chain of transmission. The first participants are trained on experimenter-designed items (termed *generation 0*); each participant's response to the input that he or she is exposed to is referred to as a *generation*, and the "genealogy" of all participant responses in the same line of descent constitutes a *chain*.

### *3.1. Methodology*

This study was conducted online with English-speaking participants. Generation 0 participants were trained on an artificial language with no consistent structure; each description contained the nouns and adpositions of the artificial language in a randomized order (for example, an array with object B atop object A could bear the description (in English) "A B atop," "atop B A," "B atop A" and so on). Participants were trained on these descriptions using the same procedure as in Experiment 1, and as in Experiment 1, in the final production stage, participants were asked to label further scenes, including both trained and novel items. The labels for trained items created in this stage were passed on as input to the next participant in the chain, using training and comprehension trials in the same format as in Experiment 1. Consequently, at each generation, participants were trained on description for the same scenes, and then asked to label novel scenes; these novel labels made it possible to observe how the grammar was being generalized and whether it was stabilizing. We ran twelve chains, each of five generations (n = 60).

### *3.2. Analysis and results*

Each string generated by a participant was compared to corresponding strings (i.e. referring to the same configuration of objects) generated by all 6 possible grammars (branching, crossed, center-embedding; head-initial or head-final), and the best match determined in the same manner as with the Experiment 1 results. We then calculated the entropy of the resulting grammar distribution (the distribution of best fit grammars for all strings produced by a given participant).

Grammar entropy of 0 bits would indicate a participant whose productions best matched those of a single grammar; entropy of 1 bit would indicate that a participant's productions were split over 2 grammars.

Grammars clearly stabilized over generations: a significant decrease in entropy occurred from 0 to 5 as measured by linear regression ($B = -.207$, $SE = .033$, $p < .001$), consistent with an increase in uniformity as irregularities were eliminated and a consistent grammar came to predominate. Figure 4 shows the best fit grammar for every production for every participant. While head-initial branching grammars (i.e. English-like grammars) dominate, one chain converges on a head-final branching grammar (chain A) and one chain (chain K) appears to converge on a crossed-dependency grammar. Center-embedding patterns did not emerge.

## 4. General discussion and conclusion

The results of these experiments are consistent with long-standing psycholinguistic findings (e.g. Bach et al. 1982) concerning the difficulty of center-embedded dependencies compared to branching structures. The status of crossed dependencies, however, remains somewhat anomalous: this form of dependency appears to be at least as easy to learn as center-embedding and possibly easier, but is very rare in natural languages. In our experiments, (strings consistent with) crossed dependency grammar emerged both as reanalyses of center-embedded input, and as a result of iterated learning (in one chain) from initially unstructured input.

In sum, these results provide additional evidence for the somewhat unexpected observation that the crossed dependency structure rarely seen in natural languages is as learnable as, or even more learnable than, the widely-found center-embedded structure. The emergence of a crossed-dependency grammar in one iterated learning chain is also interesting, especially in light of the fact that center-embedding grammar was not observed to emerge in any chain. This strikingly conflicts with the observed patterns in natural languages. These results are also significant in light of the fact that this experiment featured meaningful semantic strings, demonstrating that the anomaly of crossed and center-embedded dependencies can arise in a meaningful language-like system as well as in non-semantic symbol sequences (e.g. DeVries et al 2008, Uddén et al 2012). Why natural language typology, where grammars allowing center-embedding are common, seems to be inconsistent with learnability remains to be explained.

**Acknowledgments**

This research received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant 681942, held by K. Smith).

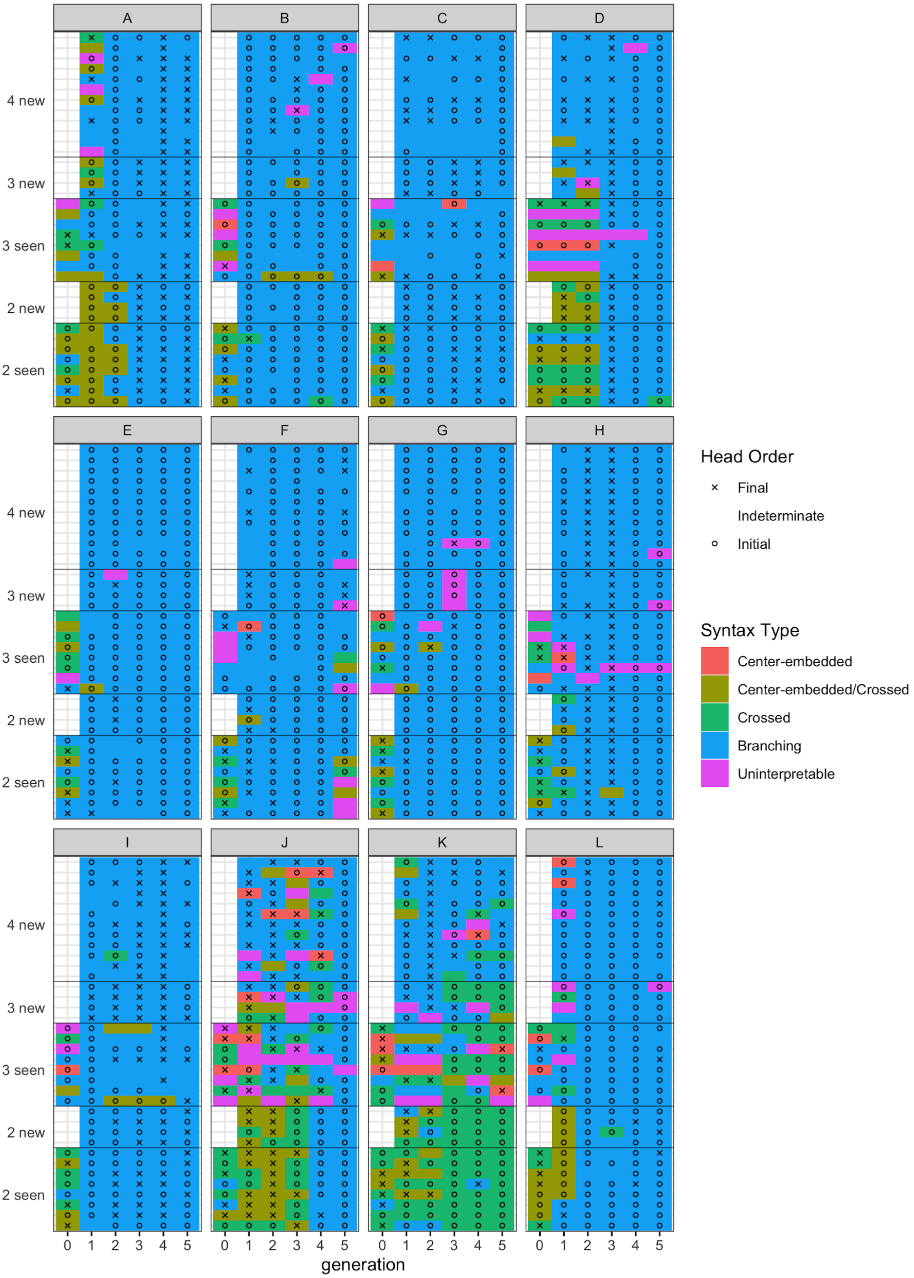


Figure 4: grammar by string, chain, and generation, Experiment 2. Each facet (A to L) is one chain; generations (participants) are columns and rows are individual scenes; the color of a given cell represents dependency type and O/X symbols represent head order, as listed in the sidebar.

## References

Bach, E., Brown, C. & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: a psycholinguistic study. *Language and Cognitive Processes, 1*, 249-262.

Bresnan, J., Kaplan, R. M., Peters, S., & Zaenen, A. (1982). Cross-serial dependencies in Dutch. In W. Savitch et al. (Eds.) *The formal complexity of natural language* (pp. 286-319). Springer, Dordrecht.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition, 122*(3), 306-329.

de Vries, M., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition, 106*, 763-774

Fedzechkina, M., Chu, B., & Florian Jaeger, T. (2018). Human information processing shapes language change. *Psychological Science, 29*(1), 72-82.

Hawkins, J. (2004). *Efficiency and complexity in grammars.* Oxford University Press.

Horst, J. S., & Hout, M. C. (2015). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48*, 1393-1409.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151–195.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681-10686.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology, 28,* 108-114.

Kuno, S. (1974). The position of relative clauses and conjunctions. *Linguistic Inquiry, 5*(1), 117-136.

Öttl, B., Jäger, G., & Kaup, B. (2015). Does formal complexity reflect cognitive complexity? Investigating aspects of the Chomsky hierarchy in an artificial language learning study. *PloS one, 10*(4), e0123059.

Uddén J, Ingvar M, Hagoort P, Petersson KM. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies:Investigating the push-down stack model. *Cognitive Science, 36* ,1078–101.

Vosse, T. & Kempen, G. (1991). A hybrid model of human sentence processing: parsing right branching, center-embedded and cross-serial dependencies. In M. Tomita and M. Kempen (Eds.) *Proceedings of the Second International Workshop on Parsing Technologies*.

# HYSTERESIS IN LANGUAGE EMERGENCE AND EVOLUTION

BART DE BOER*[1]

*Corresponding Author: bart@ai.vub.ac.be
[1]AI-lab, Vrije Universiteit Brussel, Brussels, Belgium

Correlational studies have shown that the social structure can predict properties of language – first established for population size and morphological complexity (Lupyan & Dale 2010). In emerging sign languages, different social factors appear to result in different linguistic properties. For instance, Israeli Sign language (a Deaf community sign language) shows more conventionalization and less variability than Al Sayyid Bedouin Sign Language (an equally old shared sign language) (Meir et al. 2012). Several explanations have been proposed (and supported with computer models) based on, for instance, differences between adult and infant learning[1] (Dale & Lupyan 2012) memory limitations (Thompson, Raviv & Kirby 2020) or shared cultural knowledge (Mudd, de Vos & de Boer 2022).

None of these proposed explanations are mutually exclusive, and all of them may play a role in explaining the observed phenomena. However, they all provide a one-to-one mapping between the social property (e.g. group size or social network structure) and the linguistic property (e.g. morphological complexity or lexical variability). In other words, they propose that for any setting of the social property, there is *one* optimal value of the linguistic property. This abstract presents a first exploration of the consequences of abandoning this assumption, and an argument that this may help explain real linguistic phenomena.

The existence of multiple linguistic optima for some values of the social property creates the possibility of *hysteresis* (see left panel of Fig. 1). When the social property (for instance group size) increases for a particular population (in particular during language emergence) the (linguistic) property of its language stays stuck in the lower branch of the curve. When at a later stage the social property returns to an earlier value, the language may stay stuck on the upper branch, and thus appear quite different. This may explain differences between emergent languages and established languages with similar social properties.

---

[1] Dale & Lupyan's model focuses on morphology; the other models focus more on lexical variability.

A simple model of this phenomenon can be based on statistical physics: it is assumed that the distribution of the linguistic property (e.g. complexity of morphemes) follows a Maximum Entropy Distribution (MED). Properies of systems undergoing random fluctuations will tend to have MEDs. A MED can be characterized by setting its statistical moments to have fixed values[2]. In order to allow for hysteresis, the MED must have two peaks. A possible simple model constrains the mean of the distribution and the moment given by $\int dx \cdot p(x)(2x^2 - x^4)$. The precise shape of this second constraint is unimportant, only that it has two peaks, indicating that a language makes a compromise between two pressures. The distribution (with a scaling factor Z to ensure a total probability of 1) then becomes:

$$p(x) = e^{-\alpha x + \beta(2x^2 - x^4)}/Z \tag{1}$$

where $\alpha$ represents the social property, and $\beta$ determines how high the peaks are.

Simulating a varying value of the social property with a Markov Chain Monte Carlo approach (details in the supplementary material) shows hysteresis (right panel of Fig. 1). When the property increases, transition between the optimal states happens at a higher value than when it decreases. The transition is gradual: during the transition, the language is a mixture of the two linguistic optima.

This preliminary investigation shows that hysteresis can emerge from a simple model. The challenge is now to interpret the constraints on the moments in terms of cognitive and social factors and to define empirical tests to establish whether a model like this describes the linguistic reality accurately.

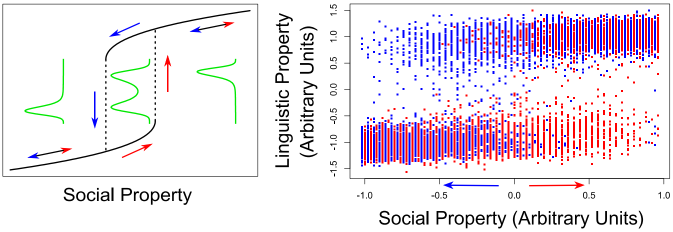


Figure 1. The left panel illustrates hysteresis qualitatively; solid black lines are trajectories of gradual change, dotted black lines jumps, and arrows indicate the direction of change. Green lines show the distribution of the linguistic property. The right panel shows hysteresis in a simulated language with 100 "morphemes" when a social property first slowly increases (red dots) and then decreases (blue dots). The transition does not occur in the same place when going up as when going down.

## Acknowledgements

This work was funded by the FWO project number G034720N and by the Flemish AI plan.

---

[2] E.g. fixing the mean gives Boltzmann's distribution; fixing variance gives the normal distribution.

## References

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems, 15*(03n04), 1150017.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS One, 5*(1), e8559.

Meir, I., Israel, A., Sandler, W., Padden, C. A., & Aronoff, M. (2012). The influence of community on language structure: Evidence from two young sign languages. *Linguistic Variation, 12*(2), 247–291. doi: 10.1075/lv.12.2.04mei

Mudd, K., de Vos, C., & de Boer, B. (2022). Shared Context Facilitates Lexical Variation in Sign Language Emergence. *Languages*, *7*(1). https://doi.org/10.3390/languages7010031

Thompson, B., Raviv, L., & Kirby, S. (2020). Complexity can be maintained in small populations: A model of lexical variability in emerging sign languages. In *Ravignani, A.; Barbieri, C.; Flaherty, M.(ed.), The Evolution of Language: Proceedings of the 13th International Conference (Evolang13)* (pp. 440-442).

# BODY SIZE PREDICTS VOCAL TRACT SIZE IN A MAMMALIAN VOCAL LEARNER

KOEN DE REUS*[1,2,3], DARYLL CARLSON[3,4], ALICE LOWRY[3,5], STEPHANIE GROSS[6], MAXIME GARCIA[7,8], ANA RUBIO-GARCIA[3], ANNA SALAZAR-CASALS[3], and ANDREA RAVIGNANI[1,3,9]

*Corresponding Author: koen.dereus@mpi.nl
[1]Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
[2]Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium
[3]Research Department, Sealcentre Pieterburen, Pieterburen, Netherlands
[4]Department of Earth System Science, Stanford University, Stanford, USA
[5]School of Environmental Science, University of Liverpool, Liverpool, UK
[6]Institute for Terrestrial and Aquatic Wildlife Research (ITAW), University of Veterinary Medicine Hannover, Foundation, Büsum, Germany
[7]Animal Behaviour, Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zürich, Switzerland
[8]Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zürich, Switzerland
[9]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark

## 1. Mechanisms for escaping acoustic allometry

Animals whose call features do not scale with their body size are said to escape acoustic allometry. These animals may thus sound smaller or larger than they are. Recent work (Garcia & Ravignani, 2020; Ravignani & Garcia 2021) found that they can achieve this by evolving vocal tract modifications (e.g., laryngeal descent; Reby & McComb, 2003) or by learning to better control their vocal organs (i.e., vocal learning, Janik & Slater, 1997). To identify which mechanism is used by species to escape acoustic allometry, one could perform an anatomical study to test if vocal tract size scales with body size. Acoustic allometry is escaped by anatomical adaptations if vocal tract size does not scale with body size, and through vocal learning if there is scaling between vocal tract size and body size.

## 2. Anatomical studies can help find more vocal learners

Studies that test whether animal vocal tracts scale with their body size offer a simple way of testing the hypothesis pitting anatomical adaptations vs. vocal learning (Garcia & Ravignani, 2020; Ravignani & Garcia 2021). Moreover, they can help to identify new species capable of vocal learning, a prerequisite for human speech. Adopting a comparative approach which considers an increasing number of vocal learning species could offer promising insights into the biological underpinnings of communication systems such as spoken language.

## 3. Harbor seal vocal tracts scale with their body size

Harbor seals' (*Phoca vitulina*) large vocal plasticity allows them to modulate the call frequencies they produce (Ralls et al., 1985; Torres Borda et al., 2021), enabling them to produce sounds with different frequencies than predicted from their body size. The current study tests if the vocal tract of the harbor seal, a known mammalian vocal learner, scales with its body size (de Reus et al., 2022). Vocal tracts, including larynges, of 68 young harbor seals (pups and weaners) were measured using a caliper and tested for allometry with body size using generalized linear mixed models. We find that both body length and body mass predict vocal tract length, vocal fold length, and tracheal dimensions (Table 1). Interestingly, allometry between body size and vocal fold length emerges after puppyhood, suggesting that ontogeny may modulate the anatomy vs. learning distinction. We conclude that the vocal tracts of harbor seals do indeed scale with body size. Consequently, by exclusion, vocal learning is the likely mechanism used in young harbor seals to escape acoustic allometry.

Table 1. Selected models for each vocal tract measurement.

| *Measurement* | *Model* | *Deviance explained* |
|---|---|---|
| Vocal tract length | BL+BM+A+S+BL*S | 59.30% |
| Vocal fold length | BL+BM+A+S+A*BM+A*S | 74.89% |
| Subglottic-tracheal dorsoventral distance 1 | BL+BM+A+S | 69.99% |
| Subglottic-tracheal dorsoventral distance 2 | BL+BM+A+S | 58.38% |

Note. BL = body length, BM = body mass, A = age class, S = sex. Predictor terms joined by an asterisk denote an interaction. Significant predictor terms are underlined.

## Acknowledgements

We would like to thank all members of Sealcentre Pieterburen for their continued support in this research. In particular, we are grateful to Letty Stupers, Rebecca Andreini, and Maria Jose Robles for help in data collection. We thank Aline Joustra for her illustrations of harbor seal vocal tracts. We also greatly appreciate the extensive comments provided by Taylor Hersh, Laura Verga, Yannick Jadoul, and Limor Raviv. This work was supported by a Max-Planck-Gesellschaft (Max Planck Research Group, MPRG) and a Fonds Wetenschappelijk Onderzoek (FWO) Pegasus Marie Curie fellowship 12N5517N awarded to A.R. K.d.R. was supported by the Fonds Wetenschappelijk Onderzoek project 'Interactive vocal rhythms' G034720N awarded to Bart de Boer. The German Ministry of Energy, Agriculture, the Environment, Nature and Digitalization (MELUND) funded the marine mammal stranding scheme and the necropsies in Schleswig-Holstein. The Sealcentre Pieterburen funded the seal stranding scheme and the necropsies at the Sealcentre Pieterburen. M.G. was supported by a research grant from the Ethologische Gesellschaft e.V., as well as a postdoctoral research grant from the Forschungskredit of the University of Zurich [grant no. FK-21-121].

## References

de Reus, K., Carlson, D., Lowry, A., Gross, S., Garcia, M., Rubio-Garcia, A., ... & Ravignani, A. (2022). Vocal tract allometry in a mammalian vocal learner. *Journal of Experimental Biology*, *225*(8), jeb243766.

Garcia, M., & Ravignani, A. (2020). Acoustic allometry and vocal learning in mammals. *Biology Letters*, *16*(7), 20200081.

Janik, V. M., & Slater, P. J. (1997). Vocal learning in mammals. In: P. Slater, C. Snowdon, J. Rosenblatt, & M. Milinski (Eds.), *Advances in the Study of Behaviour*, *26*, 59-100.

Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, Phoca vitulina. *Canadian Journal of Zoology*, *63*(5), 1050–1056.

Ravignani, A., & Garcia, M. (2021). A cross-species framework to identify vocal learning abilities in mammals. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200394.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530.

Torres Borda, L., Jadoul, Y., Rasilo, H., Casals, A. S., & Ravignani, A. (2021). Vocal plasticity in harbour seal pups. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 376, 20200456.

# LANGUAGE-SPECIFIC AND UNIVERSAL FACTORS BEHIND MORPHOLOGICAL SIMPLIFICATION: AN AGENT-BASED MODELLING STUDY OF ALORESE

Peter Dekker*[1], Marian Klamer[2], and Bart de Boer[1]

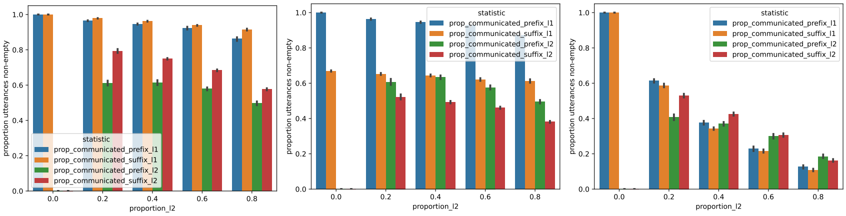
*Corresponding Author: peter.dekker@ai.vub.ac.be
[1]AI Lab, Vrije Universiteit Brussel, Brussels, Belgium
[2]LUCL, Leiden University, Leiden, The Netherlands

Large-scale correlational investigations (Lupyan & Dale, 2010) indicate that population size and the degree of contact with other groups correlate negatively with the morphological complexity of the language spoken. To study what mechanisms give rise to morphological simplification and how these surface in the world's languages, we apply agent-based models (Smith, 2014) to a case study with real-world data.

Alorese, an Austronesian language, is spoken on the coasts of the Alor and Pantar islands in Eastern Indonesia, while inward, Papuan Alor-Pantar languages are spoken. Many L1 users of Alor-Pantar languages have learned Alorese as a second language. While Alorese lost most of its morphology, its sister languages, which have not been in contact with Alor-Pantar languages, retained it (Klamer, 2012, 2020). The island geography of this case study presents us with a relatively isolated contact situation between two groups of speakers. We use agent-based models to study the hypothesis that adult language contact caused morphological simplification, focusing on inflectional verb morphology. In our model, a population of agents play a language game (Steels, 1998), in which they try to communicate concepts (verb+person) using inflected verb forms. The model is initialized with phonetic representations of verb forms in Lewoingu Lamaholot (1) (cf. Nishiyama & Kelen, 2007, p. 32), a sister language of Alorese (2) (cf. Klamer, 2011, p. 65), which lost the inflection:

(1) lodo-na Lamaholot (2) lodo-Ø Alorese
go.down-3SG go.down-3SG
'he goes down' 'he goes down'

Model outcomes are compared to Alorese, where simplification has occurred. We evaluate two cognitive mechanisms that we think could give rise to morphological simplification: the procedural-declarative model of L1/L2 language processing (Ullman, 2001) and reduction of word forms based on language-specific phonotactics. According to the procedural/declarative model, in L1 users, grammar is produced by a procedural cognitive system, while the lexicon is memorized in a declarative system. In L2 learners, linguistic forms which are normally produced in the procedural system by L1 users, such as morphology, are memorized in the declarative system. We chose to represent the procedural system in L1 users by, with a certain probability, generalizing during update: the received signal is added not only to the inferred concept, but to all concepts. In our model with phonotactic reduction, a producing agent drops the affix if adding it would lead to violation of the default CV structure of Alorese. The chosen data representation, staying close to the real word forms in Alorese, allows for this modelling of language-specific phonotactics. It should be noted that the influence of the phonotactic mechanism is more local than that of the procedural/declarative model. Our results show that the standard model (Fig 1a) does not show a substantive decrease in morphological complexity, when the proportion of L2 users increases. The phonotactic reduction model (Fig 1b), which does allow an extra possibility for simplification, does not show this decrease either. The model with generalization for L1 (Fig 1c) shows a decrease in morphological complexity for higher proportions of L2, although the complexity of suffixes stays relatively high compared to the real situation in Alorese. Based on our results, it appears that generalization in the context of the procedural/declarative model may be an important mechanism in explaining the relation between language contact and simplification. In the future, similar data-based agent-based models could be applied to case studies from other languages in the world, to identify which mechanisms are universal and which are dependent on language-specific context.



(a) Basic model. (b) Basic + phonotactic. (c) Basic + generalization L1

Figure 1.: Morphological complexity (measured by the proportion of utterances for which the affix is not empty) for different proportions of L2 users. Morphological complexity for prefix and suffix for L1 (blue/orange) and L2 (green/red).

## Acknowledgements

This work was supported by funding from the Flemish Government under the *Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen* programme. PD was supported by a PhD Fellowship fundamental research (11A2821N) of the Research Foundation – Flanders (FWO).

## References

Klamer, M. (2011). *A short grammar of Alorese (Austronesian).* Muenchen: LINCOM Europa. (OCLC: ocn760922643)

Klamer, M. (2012). Papuan-Austronesian language contact: Alorese from an areal perspective. *Language Documentation & Conservation*, *Melanesian Languages on the Edge of Asia: Challenges for the 21st Century. Special Publication No. 5*, 72–108.

Klamer, M. (2020). From Lamaholot to Alorese: Morphological loss in adult language contact. In D. Gil & A. Schapper (Eds.), *Austronesian Undressed: How and why languages become isolating* (pp. 339–367). Amsterdam: John Benjamins Publishing Company.

Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure. *PLoS ONE*, *5*(1), e8559.

Nishiyama, K., & Kelen, H. (2007). *A grammar of Lamaholot, Eastern Indonesia.* Lincom Europa.

Reali, F., Chater, N., & Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1871), 20172586.

Smith, A. D. (2014). Models of language evolution and change. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(3), 281–293. (Publisher: Wiley Online Library)

Steels, L. (1998). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In *Approaches to the Evolution of Language* (pp. 384–404). Cambridge University Press.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(2), 105–122.

# CONVERGENT CULTURAL EVOLUTION OF CONTINUERS (*mmhm*)

MARK DINGEMANSE*[1,2], ANDREAS LIESENFELD[1], and MARIEKE WOENSDREGT[2]

*Corresponding Author: mark.dingemanse@ru.nl
[1]Center for Language Studies, Radboud University
[2]Max Planck Institute for Psycholinguistics

Continuers —words like *mm, mmhm, uhum* and the like— are among the most frequent types of responses in conversation. They play a key role in joint action coordination by showing positive evidence of understanding and scaffolding narrative delivery. Here we investigate the hypothesis that their functional importance along with their conversational ecology places selective pressures on their form and may lead to cross-linguistic similarities through convergent cultural evolution. We compare continuer tokens in linguistically diverse conversational corpora and find languages make available highly similar forms. We then approach the causal mechanism of convergent cultural evolution using exemplar modelling, simulating the process by which a combination of effort minimization and functional specialization may push continuers to a particular region of phonological possibility space. By combining comparative linguistics and computational modelling we shed new light on the question of how language structure is shaped by and for social interaction.

## 1. Introduction

Social interaction is characterised by people exchanging short bursts of articulatory activity organised in turns (Sacks, Schegloff, & Jefferson, 1974). Evidence that linguistic resources are organized in response to this fact is the existence of a class of responsive items variously called response tokens, backchannels, or continuers (Yngve, 1970; Jefferson, 1985). CONTINUERS are items that occur at the boundaries of turn-constructional units and that "demonstrate both that one unit has been received and that another is now awaited" (Goodwin, 1986). Common examples of continuers so defined are forms like *mm, uh huh, yeah* in English (Indo-European) or *umm, mm, ing* in Anal Naga (Tibeto-Burman) (Fig. 1).

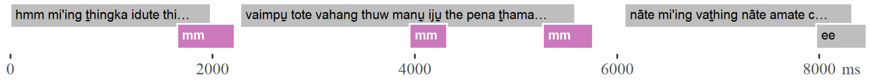


Figure 1.: Conversation in Anal Naga (Tibeto-Burman, India), showing how continuers like *mm* support another participant's production of multiple longer turns and often occur in streaks of 3+ items, a fact we use to identify candidate tokens in a language-agnostic way (data from Ozerov, 2018).

Continuers have been studied in conversation analysis and psycholinguistics (Yngve, 1970; Jefferson, 1985; Ward, 2006; Bavelas, Coates, & Johnson, 2000), but work on their cultural evolution is scarce. Yet if interaction is the primary ecology of language in use (Schegloff, 2006) and if utterances are the main target of selection in cultural evolution (Croft, 2000), then continuers, as stand-alone utterances directly exposed to the exigencies of conversation, provide a crucial locus for studying how language is shaped by interaction. We propose that key properties of items like this can be explained by thinking of them as a distinctive stratum of vocabulary (Ward, 2006) adapted to the ecology of turn-by-turn interaction.

Continuers are highly frequent words whose conversational ecology is characterized by frequent overlap and a strong skew in the amount of talk contributed by each participant (Fig. 1; cf. Rühlemann, 2018). This places special requirements on their shape (Gardner, 1997). Optimal continuers are (i) easy to plan and produce, (ii) unobtrusive, and (iii) sufficiently distinct from regular words to be seen as ceding the conversational floor. While the first criterion may be ascribed to frequency-driven effort minimization, the other two are likely to push continuers into a particular part of the possibility space. In cultural evolutionary terms, these are selective pressures that over diachronic time are likely to result in convergent solutions across unrelated languages (Dingemanse, Torreira, & Enfield, 2013).

Here we test this proposal by combining two methods. Cross-linguistic comparison provides us with primary data on the interactional ecology of continuers across languages, and can show whether indeed continuers, more than some other words, display tell-tale signs of being adapted to this ecology. Computational modelling provides us with ways to formalize and test the causal account, enabling us to see whether the proposed convergence of form can arise from cultural evolution given selective pressures of effort minimization and conversational ecology. While we focus on continuers, the results have wider relevance for language evolution in at least two ways. They point to interactional infrastructure as a key causal locus affecting the cultural evolution of linguistic items (Enfield, 2014); and they put the spotlight on metacommunication, one of the major advances in the evolution of communication (Bateson, 1972).

## 2. Methods

*Observational.* We curate transcribed corpora of natural conversation in 67 spoken languages of 28 phyla around the world (Fig. 2). Crucially, we do not search for continuers by form, but define their conversational ecology as a sequential pattern, making it possible to identify candidate tokens in a language-agnostic, non-circular way. We look for streaks of non-unique conversational turns that occur in frequent alternation with unique turns by other participants. Using a minimum streak length of 3, we identify the top 5 candidate continuer tokens per language. We also collect two contrasting sets of linguistic items for comparison. First a set we call *discontinuers*, defined as the top 5 highest frequency stand-alone turn for-

Figure 2.: The sample incorporates conversational corpora of 67 spoken languages from 28 distinct phyla, with 32 languages from 12 phyla studied in more detail.

mats that do not occur in streaks like continuers do. The interactional work done by these items is almost the converse of continuers: they do not cede the floor but take it as they redirect, resume, or repair a conversation. This means they are not subject to the same selective pressures. Second a set of *top tokens*, defined as the top 5 highest frequency tokenised words in the language (this provides a baseline comparison for frequency-driven reduction effects). Since audio processing is ongoing, most observations reported here are based on written annotations.

*Computational.* To formalise our causal account of how continuers are shaped by social interaction, we use exemplar models. Such models allow us to simulate biases affecting the shape of culturally evolving items over repeated production and perception cycles (Winter & Wedel, 2016), and have been used to study how language adapts to interaction (Roberts & Levinson, 2017). We start by reproducing a model from Wedel (2012) that aims to capture the cultural evolution of words in an abstract phonological space under various pressures. To simulate the case of continuers, we adjust the model in three ways. First, we add an overarching category distinction between a set of words that together form the 'regular vocabulary', and a set (here consisting of a single word) that constitutes the 'continuer' category. This allows us to put continuers under different pressures from regular vocabulary. Second, we increase the pressure for minimising production effort for continuers relative to regular words. Third, we remove the pressure for reuse of phonetic features for continuers. For regular words, this pressure pushes the system to make use of a limited set of phonetic features, leading to a restricted phoneme inventory, rather than a proliferation of somewhat distinct sounds. For continuers, we remove this pressure in order to allow a distinct stratum to emerge. Data and code for observational and computational methods are available through OSF: https://osf.io/v6kfn.

## 3. Results

### 3.1. *Observational*

Our language-agnostic sequential search method allows us to identify candidate continuers in all 67 languages in the sample. Figure 3 shows candidate continuers in their natural conversational ecology in 10 unrelated languages. Already these examples suggest the close intertwining of conversational habitat and continuer form. In line with their function of displaying recipiency in minimally obtrusive ways, continuers are shorter than the turns they are interspersed with, often occur in full or partial overlap, and appear to be phonologically quite minimal.

Figure 3.: Candidate continuers in 10 unrelated languages (A) shown in their natural ecology (B, annotations as in the original data), with spectrograms and pitch traces of representative tokens made using the Parselmouth interface to Praat (Jadoul et al., 2018; Boersma & Weenink, 2013) (C).

Out of the full sample there are 32 languages (of 12 phyla) with large enough corpora to yield sufficient examples of all three groups of items to be compared (we find a total of 118 continuers, 104 discontinuers, and 160 top tokens, which works out to per-language averages of 4, 3, and 5 respectively). The observations below relate to this subset of data. Because the work is ongoing and we aim for more direct comparisons of acoustic and phonetic features, we do not run inferential statistics on the preliminary observations presented here.

Top tokens are an order of magnitude more frequent than either continuers or discontinuers. They are short (2.28 characters) and relatively phonemically simple (1.8 distinct characters per token). This provides us with baseline expectations

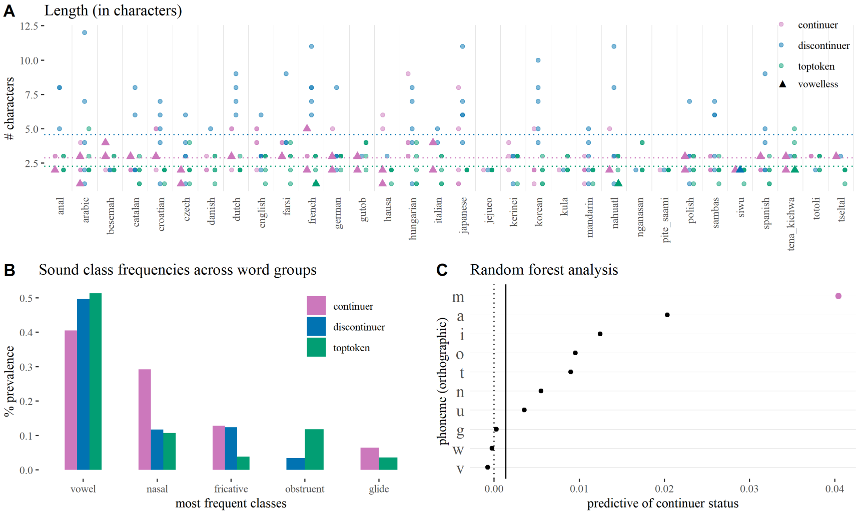


Figure 4.: **A**. Length of tokens for continuers, discontinuers and top tokens. Dotted lines mark group averages; items marked △ are vowelless. **B**. Frequencies of major sound classes across types. Vowel nuclei occur across types, but continuers stand out for their preference for nasals. **C**. Random forest analysis for 118 continuer forms in 32 spoken languages showing the top 10 most predictive phonemes (out of 29 attested). Vertical black line: absolute value of the least predictive phoneme.

for the other groups. Discontinuers are twice as long (4.1 characters) and more phonemically diverse (3.0 distinct characters per token), in line with expectations if length is affected by frequency and phonology is arbitrary. Continuers appear to be phonemically shorter than discontinuers yet longer than top tokens (2.8 characters), and as phonemically simple as top tokens (1.8 distinct characters per token).

Continuers appear to stand out in several ways. On average, one continuer token per language (32/118) is transcribed as a vowelless token like *hm, mhm, mm* (Fig. 3 and Fig. 4A), against only 1/104 discontinuers and only 3/160 top tokens. Further, 47% of continuer tokens (56/118) feature a nasal, against 32% of discontinuers (33/104)) and 24% of top tokens (39/160). Both trends are seen in a tally of phoneme classes across word types (Fig. 4B). A random forest analysis of all continuer forms together shows that ‘m’ is highly predictive of continuer status across languages (Fig. 4C). Further, 25% of continuer tokens (29/118) feature full or partial reduplication as in *mhm, hm hm, uhum, un un* (see also Fig. 3A), against only 10% of discontinuers (11/106 tokens) and 1% of top tokens (1/160).

While our main focus here has been on the 32 languages for which a comparison of the three sets is possible, the prevalence of nasals, vowelless tokens and reduplication in continuers is also seen in the remaining 35 languages, and also if we only look at 1 language per phylum (as reported in the online materials).

## 3.2. *Computational*

Figure 5 shows that whereas regular words (blue, yellow, green and orange) end up in arbitrarily different positions in the possibility space across different runs of the simulation, continuers (purple triangles) consistently end up in the same region: the centre of the graph which is the region of least effort. We also see a mild effect of this on regular words, which across independent simulation runs tend to avoid the very centre of the space as a result of the anti-ambiguity bias.

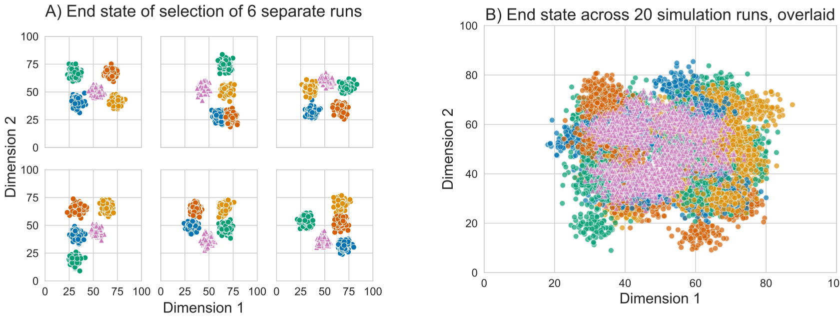


Figure 5.: Plots of where in a phonetic possibility space different words end up after 10,000 rounds of interaction, across 20 independent simulation runs (each cloud of 100 exemplar dots/triangles represents a single word at round 10,000 of a single simulation run). Blue, yellow, green and orange are regular words; purple is the continuer word. On each independent simulation run, all words are initialised at randomly selected positions in the space. **A** shows a selection of 6 separate simulation runs chosen for illustrative purposes (showing how regular words end up in different positions); **B** shows the end-state of all 20 simulation runs overlaid. Parameter settings: (i) minimal effort bias 3 times as strong for continuer word ($G$=1250) than for regular vocabulary words ($G$=5000), and (ii) the bias for reuse of features (i.e. segment-similarity bias) is not applied to the continuer category.

The simulation results show that while most words in most languages can differ arbitrarily, a distinct and cross-linguistically similar stratum of vocabulary can emerge under the cumulative effect of relatively small biases over many communicative events. They also show that commonality does not exclude diversity: while continuers are probabilistically pulled towards a similar part of the space, individual languages can still be organized in language-specific ways.

## 4. Conclusions

We have presented a first investigation into the cultural evolution of continuers. Cross-linguistic evidence suggests that spoken languages make available highly similar forms for this function. While frequency-driven reduction may partly account for the minimal shape of continuers, this is not the full story: more frequent words are still shorter on average, yet they are more phonemically diverse and less skewed in terms of word forms. Horizontal and vertical diffusion may be responsible for some similarities, but are unlikely to explain the prevalence of similar forms in over 20 independent language families. The crosslinguistic evidence is in line with the proposal that conversational infrastructure can drive convergent cultural evolution. Computational modelling supports this account by showing how a combination of effort minimization and a relaxation of the pressure for the reuse of phonetic features can push continuer-like words to the same low-effort yet distinctive part of the possibility space across languages.

The work reported here is ongoing and we are aware of a number of limitations. While written annotations are already telling, they reduce fluid signals to discrete categories, so the next step is to derive more fine-grained measures of similarity and dissimilarity from richer phonemic representations and audio signals (Fig. 3C). Further, continuers are often realized multimodally with blinks and nods; however, the subset of conversational corpora for which video is available is still limited. Finally, we have focused on spoken language, not because sign languages do not have continuers (Mesch, 2016), but because the few sign language corpora available do not systematically annotate items like this.

We have combined linguistic evidence with computational modelling to formalise and test a proposed process of conversational ecology driving convergent cultural evolution. The observational evidence suggests that continuers are pushed to a particular part of the possibility space of linguistic forms in similar ways across unrelated languages. The computational modelling shows that selective pressures enacted over cumulative cultural evolution can produce such a pattern. While modelling does not deliver definitive answers about the history of natural languages, it does shed light on the probability of explanations, and has the added virtue of requiring clarity in formulating assumptions and predictions. The findings have implications for models of cultural evolution and for our understanding of the fundamentals of human interactional infrastructure.

## Acknowledgements

We thank Andy Wedel for providing insight into his original models and Jacqueline van Arkel for help in implementing the computational model. Funding comes from Dutch Research Council (NWO) Vidi grant 016.vidi.185.205 to MD.

## References

Bateson, G. (1972). *Steps to an ecology of mind; collected essays in anthropology, psychiatry, evolution, and epistemology.* University of Chicago Press.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer.*

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach.* Harlow: Pearson Education Limited.

Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). Is ”Huh?” a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLOS ONE*, *8*(11), e78273.

Enfield, N. J. (2014). *Natural Causes of Language: Frames, biases and cultural transmission.* Berlin: Language Science Press.

Gardner, R. (1997). The Conversation Object Mm: A Weak and Variable Acknowledging Token. *Research on Language & Social Interaction*, *30*(2), 131–156.

Goodwin, C. (1986). Audience diversity, participation and interpretation. *Text*, *6*(3), 283–316.

Jadoul, Y., Thompson, B., & De Boer, B. (2018). Introducing parselmouth: A python interface to praat. *Journal of Phonetics*, *71*, 1–15.

Jefferson, G. (1985). Notes on a systematic Deployment of the Acknowledgement tokens 'Yeah' and 'Mmhm'. *Papers in Linguistics*, *17*(2), 197–216.

Mesch, J. (2016). Manual backchannel responses in signers' conversations in Swedish Sign Language. *Language & Communication*, *50*, 22–41.

Ozerov, P. (2018). *A community-driven documentation of natural discourse in Anal, an endangered Tibeto-Burman language.* (Tech. Rep.). Endangered Languages Archive.

Roberts, S. G., & Levinson, S. C. (2017). Conversation, cognition and cultural evolution: A model of the cultural evolution of word order through pressures imposed from turn taking in conversation. *Interaction Studies*, *18*(3), 404–431.

Rühlemann, C. (2018). TCU-initial backchannel overlap in storytelling. *Narrative Inquiry*, *28*(2), 257–279.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735.

Schegloff, E. A. (2006). Interaction: The Infrastructure for Social Institutions, the Natural Ecological Niche for Language, and the Arena in which Culture is Enacted. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 70–96). Oxford: Berg.

Ward, N. (2006). Non-lexical conversational sounds in American English. *Pragmatics & Cognition*, *14*, 129–182.

Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition*, *4*(4), 319–355.

Winter, B., & Wedel, A. (2016). The Co-evolution of Speech and the Lexicon: The Interaction of Functional Pressures, Redundancy, and Category Variation. *Topics in Cognitive Science*, *8*, 503–513.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting, Chicago Linguistic Society* (pp. 567–578).

# AMPLIFYING THE BIAS TO ORDER NOUNS BEFORE ADJECTIVES: STUDIES OF SILENT GESTURE AND EMERGING SIGN LANGUAGE

Monica L. Do*[1], Simon Kirby[2], Susan Goldin-Meadow[3], Natasha Abner[4], Molly Flaherty[5], Ann Senghas[6]

*Corresponding Author: monicado@uchicago.edu
[1]Linguistics, University of Chicago, Chicago, IL, USA
[2]Centre for Language Evolution, University of Edinburgh, Edinburgh, UK
[3]Psychology, University of Chicago, Chicago, IL, USA
[4]Linguistics, University of Michigan, Ann Arbor, MI, USA
[5]Psychology, Davidson College, Davidson, NC, USA
[6]Psychology, Barnard College, New York, NY, USA

Cross-linguistically, roughly 64% of the world's languages order Nouns before Adjectives, compared to the roughly 27% with the reverse order (Dryer, 2013). Yet, the factors driving this cross-linguistic tendency towards the NounAdj (i.e., post-nominal) order remains largely unknown. To investigate this asymmetry, we analyze production corpora elicited from three cohorts of Nicaraguan Sign Language (NSL1, NSL2, NSL3; Exp1a) and Silent Gesturers (Exp1b). Exp2 then examined the contribution of learning biases such as the bias towards regularization (Hudson Kam & Newport, 2005, 2009; Singleton & Newport, 2004) to the prominence of the NounAdj order cross-linguistically.

In Exp1a (NSL 1-3), we extracted 223 manual utterances containing a Noun and at least one Adjective and coded for order (NounAdj or AdjNoun) within each Noun Phrase; utterances where order was ambiguous (e.g., AdjNounAdj/NounAdjNoun) were excluded. Binomial tests showed a reliable preference for the NounAdj order for each NSL cohort (Table 1). Linear mixed models using forward difference contrasts revealed no difference between cohorts (p's > .3), suggesting that the bias towards NounAdj did not strengthen over consecutive generations.[1] In Exp1b, we analyzed 276 Noun Phrases from a production corpus of English-speakers asked to gesture without talking (Silent Gesturers). Although NounAdj preference was weaker for Gesturers compared to NSL cohorts (p < .01), we nonetheless found a preference for NounAdj among Gesturers (Table 1). Thus, manual productions from signers of an emerging

---

[1] In Exp1, we also collected utterances from Nicaraguan and Guatemalan Homesigners. However, data from these groups were too sparse to submit for analysis.

language and from native speakers of an AdjNoun language both point to a “natural” (Goldin-Meadow et al., 2008) way of ordering Nouns versus Adjs.

Table 1: Proportion of NounAdj versus AdjNoun utterances for each participant group. Raw counts are given in parentheses.

| Language Group | NounAdj | AdjNoun | SD | p | 95% CI |
|---|---|---|---|---|---|
| NSL1 (n=8) | .84 (62) | .16 (12) | .37 | < .001 | (.73, .91) |
| NSL2 (n=6) | .89 (62) | .11 (8) | .32 | < .001 | (.79, .94) |
| NSL3 (n=8) | .87 (69) | .13 (10) | .33 | < .001 | (.78, .94) |
| Silent Gesturers (n=20) | .61 (168) | .39 (108) | .49 | < .001 | (.55, .67) |

In addition to the cognitive bias towards NounAdj, we also wanted to see how other biases in language – such as the well-known bias toward regularizing variation in an input language (e.g., Hudson Kam & Newport, 2005, 2009; Singleton & Newport, 2004) – might simultaneously contribute to the prevalence of the NounAdj order cross-linguistically. Exp2, therefore, recruited 160 English-speakers to a web-based silent gesture regularization study. Comprehenders saw an event (e.g., someone waving a spotted spoon) and then were trained on two gesture vignettes describing that event. Vignettes differed only in the order of the Noun versus Adj gestures. Critically, the frequency that participants saw NounAdj versus AdjNoun vignettes varied by condition. In Majority NounAdj Conditions, they saw NounAdj and AdjNoun orders in 75% versus 25% of trials, respectively. Frequencies were reversed in the Majority AdjNoun Conditions.

We analyzed Entropy Change scores (Ferdinand et al., 2019; Shannon, 1948) and proportion of Majority Order selections using mixed models. In line with predictions, Entropy Change scores showed evidence of regularization in Majority NounAdj (p’s < .01) but not Majority AdjNoun conditions. Participants also selected Majority Order vignettes more frequently when vignettes were in the NounAdj configuration (p < .01). Thus, participants were more willing to regularize towards the NounAdj order than to the AdjNoun order.

Our results point to two factors driving the cross-linguistic prominence of NounAdj word orders. The first is a cognitive bias for NounAdj orders stemming from a “natural” way of representing objects and their attributes. The second factor is a regularization bias amplifying those underlying preferences. Ongoing work investigates the NounAdj order in silent gesture communication and in iterated learning paradigms.

## Acknowledgements

This work was funded by the NIDCD of the NIH, R01 DC005407 awarded to A. Senghas and NSF Grant #1654154 awarded to S. Goldin-Meadow.

## References

Dryer, M. S. (2013). Order of Adjective and Noun. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/87

Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, 184, 53–68. https://doi.org/10.1016/j.cognition.2018.12.002

Goldin-Meadow, S., So, W. C., Ozyurek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, 105(27), 9163–9168. https://doi.org/10.1073/pnas.0710060105

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, 1(2), 151–195. https://doi.org/10.1080/15475441.2005.9684215

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30–66. https://doi.org/10.1016/j.cogpsych.2009.01.001

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*, 49(4), 370–407. https://doi.org/10.1016/j.cogpsych.2004.05.001

# THE ROLE OF STRUCTURAL PRIMING, SEMANTICS AND POPULATION STRUCTURE IN WORD ORDER CONVENTIONALIZATION: A COMPUTATIONAL MODEL

Loïs Dona*[1] and Marieke Schouwstra[2]

*Corresponding Author: l.m.dona@students.uu.nl
[1] Information and computing sciences, Utrecht University, Utrecht, The Netherlands
[2] Humanities, University of Amsterdam, Amsterdam, The Netherlands

## 1. Population structure and word order regularization

There is an ongoing debate about the influence of population structure on the emergence of linguistic rules. Some stress the importance of *population size*, or factors often confounded with it (Raviv, Meyer, & Lev-Ari, 2019; Thompson, Raviv, & Kirby, 2020; Lupyan & Dale, 2010; Wray & Grace, 2007; Lou-Magnuson & Onnis, 2018), while others argue for the importance of *network connectivity* (the degree of diversity in agent encounters) (Segovia-Martín, Walker, Fay, & Tamariz, 2020; Richie, Yang, & Coppola, 2014). In a series of simulations, we investigate the influence of population structure on the conventionalization of basic word order. Of the languages in the world, many use a dominant order of subject, object and verb for conveying who did what to whom (Dryer, 2013). Previous research has suggested that in the early stages of language emergence, people use word orders variably, depending on event semantics: SOV order for extensional events (in which the direct object is specific and concrete; e.g. pirate throws ball), and SVO for intensional events (in which the direct object is more abstract, and possibly dependent on the verb; e.g. pirate thinks of ball) (Schouwstra & Swart, 2014). Structural priming is a potential mechanism driving increased regularity in word order (Christensen, Fusaroli, & Tylén, 2016; Schouwstra, Smith, & Kirby, 2020). A key question is how properties of the population influence regularization.

## 2. Local regularization: interaction in pairs of agents

Our first model investigates the relative contributions of semantics and structural priming over time. We compare three different simulations of dyadic interaction, one where the influence of structural priming increases over time and that of semantics decreases (*increasing-priming*), one where it is the other way around (*decreasing-priming*), and one where their influences are equal and constant (*constant-priming*). The increasing and decreasing influence of priming are modelled by an exponential function with time as the exponent. The influence of semantics is conversely modelled as $1 - priming$. The influence shifts were accomplished by treating the probability of choosing SOV or SVO word order as a linear combination of previously observed and produced word orders that were SOV or SVO (structural priming), modelled as 1 and 0 respectively, and event semantics. Regularization was measured by calculating the change in mutual information over time. A comparison of the simulations shows that only when the influence of priming increases relative to the influence of event semantics, word order becomes more regular. By contrast, when structural priming is constant over time, this does not lead to increased regularity.

## 3. Simulating population properties

To assess the effects of population properties on word order regularization, we simulated populations with different sizes and densities, copying the agent properties of the increasing priming simulation above. We measured network connectivity of these populations using the distribution of the number of observations made by every agent from every other agent after every thirty two interactions. We used the slope parameter of a curve fitted to mutual information data over time as a measure for regularization speed.

The results show a positive correlation between population density and connectivity, and a negative correlation between population size and connectivity (using Spearman's rank correlation Myers & Sirois, 2004). Both size and density correlate significantly with regularization speed (see figure 1), but density correlates significantly stronger (Fisher Z-transform.; $z = -16.74, p < 0.001$).

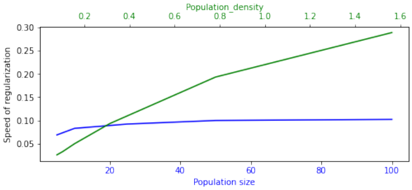


Figure 1. Population size and density plotted against regularization speed ($k$).

The relatively modest effect of population size on regularization speed indicates that population size alone is not an explanatory factor for the differences in convergence rates and linguistic simplicity between communities, but often correlates with these effects due to it being confounded with network connectivity.

These findings contribute to the unraveling of the forces at play during word order conventionalization on the cognitive and population level. Moreover, the computational structure of the model in this paper allows for easy incorporation of other potential influences (e.g. effects of language contact). Recent work has shown that viewing word order as continuous rather than discrete might be more realistic (Levshina et al., 2021), and that word order regularization is also reflected in learning (Motamedi, Wolters, Naegeli, Schouwstra, & Kirby, 2021). The current framework makes it possible to explore these new insights further.

## References

Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, *146*, 67–80.

Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Levshina, N., Namboodiripad, S., Allassonnière-Tang, M., Kramer, M. A., Talamo, L., Verkerk, A., Wilmoth, S., Rodriguez, G. G., Gupton, T., Kidd, E., et al.. (2021). Why we need a gradient approach to word order.

Lou-Magnuson, M., & Onnis, L. (2018). Social network limits language complexity. *Cognitive Science*, *42*(8), 2790–2817.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1), e8559.

Motamedi, Y., Wolters, L., Naegeli, D., Schouwstra, M., & Kirby, S. (2021). Regularisation, systematicity and naturalness in a silent gesture learning task. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Myers, L., & Sirois, M. J. (2004). Spearman correlation coefficients, differences between. *Encyclopedia of statistical sciences*, *12*.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*(1907), 20191262.

Richie, R., Yang, C., & Coppola, M. (2014). Modeling the emergence of lexicons in homesign systems. *Topics in cognitive science*, *6*(1), 183–195.

Schouwstra, M., Smith, K., & Kirby, S. (2020). The emergence of word order conventions: improvisation, interaction and transmission.

Schouwstra, M., & Swart, H. de. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436.

Segovia-Martín, J., Walker, B., Fay, N., & Tamariz, M. (2020). Network connectivity dynamics, cognitive biases, and the evolution of cultural diversity in round-robin interactive micro-societies. *Cognitive science*, *44*(7), e12852.

Thompson, B., Raviv, L., & Kirby, S. (2020). Complexity can be maintained in small populations: a model of lexical variability in emerging sign languages.

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*(3), 543–578.

# SYLLABLE SEQUENCE PERCEPTION IN SONGBIRDS AND PARROTS

ROBERT J. DOOLING*[1], HSIAO-WEI TU[1], ADAM R. FISHBEIN[1], KATHERINE A. STENNETTE[1].

*Corresponding Author: Rdooling@umd.edu
[1]Psychology Department, University of Maryland, College Park MD USA

The capacity to use recursive syntax is considered a fundamental difference between human and animal vocal communication (Hauser and Fitch, 2003). Even more fundamental, however, is the capacity to produce and perceive acoustic sequences. The sequentially complex vocalizations of songbirds have long served as a potential bridge between human language and animal vocalizations with much of the historical evidence coming from field studies (e.g., Kroodsma and Miller, 1996). Laboratory perceptual studies, on the other hand, have provided less compelling evidence that songbirds have the perceptual machinery to process complex sound sequences (Fishbein, et al, 2019; Geberzahn and Deregnaucourt, 2020; Lawson, et al, 2018; Seki, et al, 2013). By contrast, laboratory studies on one psittacine, the budgerigar (*Melopsittacus undulatus*), has revealed that this species can both produce a sequentially complex vocal repertoire consisting of multiple syllable categories and perceive changes to syllable sequence in its song (Fishbein, et al, 2019; Tu and Dooling, 2012; Tu, et al, 2011).

A major difficulty in studying sequence production and perception in birds is that cross species comparisons, in which each species is tested under identical conditions, are rarely done. Here we first showed that the complex sequential structure of the budgerigar warble song can be modeled as a 5th order Markov sequence. Then we successfully trained budgerigars and various songbirds, including canaries (*Serinus canaria*), zebra finches (*Taeniopygia guttata)*, Bengalese finches (*Lonchura striata domestica)*, and Cordon Bleus (*Uraeginthus bengalus)* by operant conditioning to discriminate changes in the sequential order of various natural vocalizations as well as neutral sounds including tones, words, and syllables within a word.

Results show that budgerigars, a psittacine, can discriminate changes in the sequence of syllables in their warble and in other sound sequences. This ability matches well with the sequential complexity of their warble. Several species of songbirds tested on these sequences of sounds, and their own species-specific vocalizations, cannot discriminate these perturbations in syllable or sound sequences. Even reducing the task to its simplest level, discriminating a change in the order of two syllables, budgerigars well outperform songbirds even when both budgerigars and songbirds are tested on songbird species-specific vocalizations.

While there is a long history of efforts aimed at understanding the differences in acoustic communication between humans and birds, the focus of much of this comparative work has been on the production side (e.g., Kroodsma and Miller, 1996). Our results suggest that the limits to acoustic communication using sound sequences in birds could lie more on perceptual side. Indeed, the differences observed here between songbirds and parrots may offer hints as to how the perception of sequences evolved in the primate lineage to support the emergence of human language. It might be fruitful to focus on anatomical differences between songbird and parrot brains to identify the neural candidates that might underly these significant differences in the perception of complex acoustic sequences.

## Acknowledgements

This work has been supported by both National Institutes of Health and the National Science Foundation.

## References

Hauser, M. D., & Fitch, W. T. (2003). What are the uniquely human components of the language faculty? In M. H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 158-181). Oxford: Oxford University Press.

Fishbein, A. R., Idsardi, W. J., Ball, G. F., Dooling, R. J. (2019) Sound sequences in birdsong: how much do birds really care? Philosophical Transactions of the Royal Society B. 2019.

Geberzahn, N. and Deregnaucourt, S. (2020). Individual vocal recognition in zebra finches relies on song syllable structure rather than song syllable order. Journal of Experimental Biology, 223, 1-9.

Lawson, S. L., Fishbein, A. R., Prior, N. H., Ball, G. F., and Dooling, R. J. (2018) “Relative salience of syllable structure and syllable order in zebra finch song,” Animal Cognition 21, 467-480.

Mol, C., Bolhuis, J.J., and Moorman, S. (2021). “Vocal learning in songbirds: the role of syllable order in song recognition”. Philosophical Transactions B. Biological Science, 376.

Seki, Y., Suzuki, K., Osawa, A. M., and Okanoya, K. (2013) “Songbirds and humans apply different strategies in a sound sequence discrimination task,” Frontiers in Psychology 4, 1-9.

Tu, H.-W. and Dooling, R.J. (2012) Perception of warble song in budgerigars (Melopsittacus undulatus): evidence for special processing. Animal Cognition, 15(6), 1151-1159.

Tu, H.-W., Smith, E.W., & Dooling, R.J. (2011) Acoustic and Perceptual Categories of Vocal Elements in the Warble Song of Budgerigars (Melopsittacus undulates). Journal of Comparative Psychology, 125(4), 420-30.

# NEURONAL RECYCLING AS SPEECH ORIGINS

AXEL G. EKSTRÖM*[1]

*Corresponding Author: axeleks@kth.se
[1]KTH Speech, Music & Hearing, Stockholm, Sweden

The neuronal recycling hypothesis posits that cultural inventions invade evolutionarily adapted neural circuitry and facilitate predictable neurological development in humans. I present an account of phonological development as derived from that same set of proposed principles. So applied, the hypothesis predicts (1) observable neural biases and abilities of communicative learning continuous across primates, (2) patterns of cross–culturally expectable phonological development in human infants, and (3) principled manifestation of phonologies across human cultures. I argue that all three are supported by empirical work in primatology, neurolinguistics, developmental psychology, and phonetics. Speech is invented anew in every speaking child, but representations of speech-centric behavior are consistently allocated to the same neural architecture across individuals – structures that house related functionality in non–human primates. Development of speech production and perception also follow predictable trajectories across cultures. Relevant scientific findings thus support the view that human infants are equipped with neural architecture evolutionarily prepared for the development of vocal–communicative behavior, which skew their ontogenetic development and constrain their ultimate expression. Implications for speech ontogenesis are discussed. The present work adds to the understanding of systems of speech as products of cultural evolution.

## 1. Introduction

All humans who have ever held a conversation have solved the same problem in ontogeny. They have learned to speak. Yet, despite their importance, the mechanisms by which speech is learned by the growing child remain poorly understood, and in urgent need of hypothesizing. In this work, I address the ontogenetic development of speech, and the evolutionary roots from which learned experience drives emergent speech behavior. I am not here concerned with aspects of social cognition (e.g., intentionality), but the biological architecture that underlies spoken language in particular. A view of speech from evolution must base its assumptions on principles of the natural world, without resorting to abstract theorizing. I address the fundamental question of how abilities of speech may develop in the individual, without taking out significant loans that theories of evolution cannot hope to cash. The present stated goal, rather, is an account of speech ontogenesis that derives speech, and its neural, and perceptual features, from more basic elements of non–speech behavior.

The process by which phylogenetically adaptive neural structures are coopted for new uses has been described by various researchers (see Andersson, 2010). For one, Dehaene (2009) has argued that preexisting evolutionarily adaptive architectures in the human brain enact constraints upon the development of writing systems – cultural inventions too recent to have driven any significant evolutionary selection pressure – such that they exhibit principally emergent features across cultures. Accordingly, reading in the brain predictably occupies the same neural regions across individuals (i.e., the visual word form area). This set of theoretical assumptions is referred to as the neuronal recycling hypothesis (hereafter NR). By analogy with reading, I posit that speech development is similarly non-arbitrary but follows a predictable protocol. It is not claimed that speech as such is solely a learned behavior; rather, speech capacities likely represent a relatively recent evolutionary heritage in *Homo* (see Ghazanfar et al., 2012; MacNeilage, 1998; Lieberman, 2017; Lieberman et al., 2019). However, speech is also invented anew by every developing child, and the processing and production of speech is localized to the same neural architecture across individuals, languages, and cultures, but selectively sensitive to language-specific phonologies and linguistic cues. As cultural mutually agreed–upon conventions, human spoken languages possess multiple distinct common features, facilitating formalization of linguistic codes and usage. The NR framework helps explain several puzzling findings in speech-centric science, including observed constraints on individual linguistic development in ontogeny, and the apparently limited size of phonological repertoires inherent to particular languages. The text thus also places human phonology in a cultural evolution framework, such that systems of speech, themselves products of culture, exhibit influence over human linguistic development. The present account argues that the fallout of such organization in turn tunes and organizes the neural circuitry of the infant and child exposed to it.

## 2. Neuronal recycling for speech

Applied to speech, NR can be postulated as a set of three hypotheses (Dehaene, 2009; Christiansen & Müller, 2015). In the follow sections, support for each hypothesis is discussed at length. First, NR predicts (1) phylogenetic continuity across related species. For speech, such a supposition touches on a rich and growing research tradition on the linguistic capacities of non–human primates (hereafter primates). Further, NR anticipates (2) predictable linguistic development in infants across cultures. By now, the trajectory of children's acquisition of speech capacities has been described in significant detail, allowing for a composite illustration. Finally, NR posits (3) limited variability in expression, resulting from the predispositions of adapted neurological structures toward learning prelusive behavior. Systems of speech are indeed characterized by a range of cross-cultural commonalities. Regardless of language, culture, and geography, speech-centric behavior is allocated to the same neural architecture but develops selective sensitivities to language-specific phonological elements.

### *2.1 Primate speechlike behavior*

Articulate speech is a novel invention in the hominid lineage, its biomechanical components having evolved gradually. Because humans are the only (extant) species capable of generative speech, a proper cross–species continuity may initially appear impossible, or even undesirable. However, recent research on the communicative abilities and vocal capacities of primates allows for careful inference with relevance for the evolution of speech. Because language does not fossilize, bioanthropological research has instead focused on identifying biological precursors to speech. While the anatomy of the vocal tract itself is highly conserved in mammalian species (Negus, 1949) and thus unlikely to have contributed to the evolution of speech, much discussion has centered on uniquely human adult anatomy likely to enable speech (e.g., Lieberman et al., 2001; Lieberman, 2017). Of potentially greater relevance to understanding ontogenetic emergence of phonology, however, are studies of vocalizations themselves.

While primate communicative repertoires often include either vocalization or jaw–driven speech–like motor behavior, such as lip–smacking (e.g., Ghazanfar et al., 2012), the ability to couple the two is typically absent. Further, the acoustic structure of primate calls appears largely fixed, such that call behaviors develop reliably regardless of social experience (e.g., Cheney et al., 1992). A proper discussion of anatomical and neural limitations precluding primates from human or humanlike speech is beyond the scope of the present text and remains the subject of some controversy (see e.g., Lieberman, 2017). As a behavioral composite of voluntary mandibular and laryngeal control, speech is contingent on cortical coupling (i.e., overlap) between jaw and larynx muscle primary motor somatotopic representations (MacNeilage, 1998; Brown et al., 2021). In humans, the laryngeal motor cortex – the primary cortical center for control of vocalization – is located in the ventral primary motor cortex; in primates, the same region is located in the ventral premotor cortex. Thus, the primary neural region of vocalization control has undergone a dorsal shift with the evolution of the *Homo* genus, likely contributing to the evolution of speech (Simonyan & Horwitz, 2011). Further, comparative neuroscience suggests that neural architecture commonly associated with human speech have correlates in the brains of other hominids, with species such as Chimpanzees (*Pan troglodytes*) possessing neuroanatomical homologues to Broca's area (Brodmann areas 44, 45), which appears to activate during vocal signaling (see e.g., Taglialatela et al., 2011).

Finally, there is substantial continuity in vocal perception across the primate lineage. Research on the call repertoires of primates suggests that modification of produced calls remains relatively limited across the lifespan, with more flexibility in perception than production (e.g., Cheney et al., 1992). Such findings are echoed by observations of human infants, who develop the ability to perceive and differentiate ingroup language-specific phonemes around the age of six months – long before they begin producing it reliably and willingly. Neurologically, speech perception and language–specific perception in the Broca's area also develop during the first year of life (Imada et al., 2006), whereas the articulatory vocal production organs take several months yet to bring under voluntary control.

### *2.2. A view from ontogeny*

In human infants, mandibular oscillatory motor movements later coopted for syllable production are exclusively utilized for voiceless sucking (i.e., feeding) behavior, and are not coupled with motor movement of the vocal tract. Vocalizations observed during this stage include crying and are – unlike speech proper – innate, involuntary, and likely contingent on the same neural circuitry that underlies innate call repertoires observed in other hominids (Jürgens, 2009). Acoustically, infants' crying aligns more closely with primate vocalizations, than with adult human speech (Lieberman et al., 2019). Indeed, stemming from anatomical limitations, the infant vocal apparatus is likely incapable of producing anything resembling the range of speech sounds available to the adult speaker. Only around the age of six months does this begin to change, with the pre–verbal infant's invention of babbling, a form of emergent protospeech characterized by articulate non–word speech sounds (i.e., 'vocal play'; see Vihman, 2013). While there has been significant debate over the relationship of babbling to speech proper, researchers have long since mapped its developmental trajectory.

Human infants typically begin to babble by no later than six months of age. Around the same time, infants also begin to display an ability to discriminate perceptually between different vowel and consonant sounds (see Oller, 2000). At this point in development, infants exhibit the emergent ability to voluntarily open and close the vocal tracts. NR further predicts the observed development of speech–centric behavior in the growing infant (for an overview, see Vihman, 2013). The innate mandibular–oscillatory motor behavior utilized for suckling provide a scaffold for the earliest occurring vowel-like "cooing" sounds, which are predictably produced by human infants in the first few months (Oller, 2000). While languages exhibit significant variability in vowel systems (Maddieson, 1984), such early pseudo vowels roughly correspond to the adult /a/ and /u/, which appear almost universally across languages. Similarly, phonemes composed from bilabial plosives such as /maː/ or /baː/ are more readily produced and acquired by the individual infant early in development, around the age of 10 months. It is thus not surprising that even unrelated languages have similar–sounding words for 'mother' – tellingly, some variation of /maː/ or /məˈmɑː/. This stands in contrast to, for example, phonemes composed from fricative alveolars such as /taː/ or /kaː/, which require the use of the tongue tip and tongue body, respectively; substantial articulatory maneuverability and muscle dexterity need develop before the production of such sounds becomes possible. That is, a developmental threshold must be passed before relatively complex speech sounds can be achieved.

Neurologically, then, at this stage of development, the biomechanical equipment necessary for organizing speech into consonant 'frames', and consonant 'content' (MacNeilage, 1998) has likely begun to take place in the infant; however, the process of adapting to the ambient soundscape and local phonology (i.e., available phonemes inherent to the language spoken by present ingroup members; Goldstein et al., 2003) is not yet completed. Rather, the infant has reached a stage of orosensory exploration, where possible speech sounds – made available by combinations of laryngeal and mandibular motor activity – can be mapped onto perceptual phonemes (Kuhl, 2000). The resulting composite emphasizes their relationship in perceptual memory, enabling successful encoding across time. Thus, while a set of phonemes may be 'given', infants are born with a great range of possible developments, which are subsequently pruned through cultural learning and ingroup interaction. Such plasticity is greater in children than adult humans, facilitating acquisition of speech-centric behavior. Thus, in adult language learners, foreign speech sounds. For example, the Japanese language lacks the phoneme /l/ as observed in e.g., English, and as a result, adult Japanese speakers readily experience difficulty perceiving and producing it (see Yamada & Tohkura, 1992). Crucially, while being able to reliably discriminate /l/ from /r/ at six months of age, at eight months, Japanese infants have apparently lost this ability (Tsuhima et al., 1994). That is, cultural learning via attention to, and interaction with, adult caregivers have formed the basis for later-in-life language-specific phonological speech production (Goldstein et al., 2003). That is, much of the necessary recycling takes place in the early years of development, after which much of the plasticity required for achieving articulatory fluency with regard to a given phonology is exhausted.

### *2.3. Principled manifestation*

NR predicts constraints on cultural–environmental factors, such that culture influences the development of evolutionarily adaptive brain circuits – but not unlimitedly so. All languages are indeed organized through similar constraints on composition (e.g., grammar; but for a nuanced discussion, see Christiansen & Chater, 2008) regulating their combinatorial capacities; and content (i.e., words as units of transmission). Additionally, in another homologue to writing systems, representations of speech–centric behavior and processing is also predictably allocated to the same neural architectures (e.g., Brodmann areas 22, 44, and 45) in all human brains. Yet, human infants are born without language, and need to develop substantially before language acquisition becomes cognitively feasible.

As culturally agreed-upon systems of communication, languages exhibit significant cross-cultural similarities. Whereas all (spoken) languages constitute phonologically combinatorial systems, attempts have been made to formalize their structure according to phonological universals. The most successful of these posits that spoken language is typically organized in consonant/vowel cycles (MacNeilage, 1998). Further, spoken languages contain a limited pool of speech sounds. While phonemic libraries vary significantly between languages (Maddieson, 1984), the organization and character of those libraries are naturally constrained by the mechanical apparatus of speech. Namely, the human articulators – the tongue, upper and lower lips, alveolar ridge, hard palate, velum, uvula, pharyngeal wall, and glottis – allow for a large though ultimately limited set of articulate speech sounds. The resulting phonemic repertoire, averaging ~40 phonemes in a given language (Maddieson, 1984), however, is constrained in not only the dimension of production – resulting from biological and anatomical limitations of the vocal cords, vocal tract, and phonemic articulators – but also that of perception, operationalized as limitations on cognitive processing and memory, with classic work on optimal maximization of vowel space by Liljencrants and Lindblom (1972) indicating a crucial role of perceptual contrast in the emergent organization of phonetic structure of language vowel systems. Relatedly, Christianson and Chater (2008) have argued for a set of cognitive constraints on language evolution, shaped by limitations of the human brain. The particular repertoire of any given spoken language should thus be seen as resulting from a combination of economic-organizational and articulatory limiting factors. Whereas the most prominent limitations on phonemic expression are biological and anatomical, principles of economy of memory impose further limitations on the emergent phonemic repertoire recruited into systems of speech.

## 3. Conclusions

I have presented an account of speech ontogenesis, based on principles of neurobiology, phonetics, and observations of phonological and linguistic development. It is the view presented here that systems of speech, as learned cultural practices enact significant pressure on phonological and neurological development. While the text is primarily concerned with evolutionary, developmental, and neurological aspects of phonological activity, it aligns with a cultural evolution framework. Exposure to language-specific phonology prunes the infant listener for continued interaction and future speech sound production and perception. Speech systems – motor as well as orosensory – are subject to extensive such pruning throughout infancy and early childhood in preparation and accommodation for native-language speech sounds. Language-specific speech behavior can thus be seen as a product of recycling by cultural systems of speech.

## References

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences*, *33*(4), 245-266.

Brown, S., Yuan, Y., & Belyk, M. (2021). Evolution of the speech-ready brain: The voice/jaw connection in the human motor cortex. *Journal of Comparative Neurology*, *529*(5), 1018‑1028.

Cheney, D. L., Owren, M. J., Dieter, J. A., & Seyfarth, R. M. (1992). 'Food'calls produced by adult female rhesus (Macaca mulatta) and Japanese (M. fuscata) macaques, their normally‑raised offspring, and offspring cross‑fostered between species. *Behaviour*, *120*(3‑4), 218‑231.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, *31*(5), 489-509.

Christiansen, M. H., & Müller, R. A. (2015). Cultural recycling of neural substrates during language evolution and development. *The cognitive neurosciences V*, 675‑682.

Dehaene, S. (2009). Reading in the brain. *New York*.

Ghazanfar, A. A., Takahashi, D. Y., Mathur, N., & Fitch, W. T. (2012). Cineradiography of monkey lip–smacking reveals putative precursors of speech dynamics. *Current Biology*, *22*(13), 1176–1182.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport*, *17*(10), 957–962.

Jürgens, U. (2009). The neural control of vocalization in mammals: a review. *Journal of Voice*, *23*(1), 1-10.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850-11857.

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of oral biology*, *46*(2), 117–128.

Lieberman, P. (2017). Comment on "Monkey vocal tracts are speech-ready". *Science Advances*, *3*(7), e1700442.

Lieberman, P., Harris, K. S., Wolff, P., & Russell, L. H. (2019). *Newborn infant cry and nonhuman–primate vocalizations* (pp. 41–53). De Gruyter Mouton.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 839-862.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and brain sciences*, *21*(4), 499–511.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Negus, V. (1949). *The comparative anatomy and physiology of the larynx.* Heinemann.

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist*, *17*(2), 197–208.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PloS one*, *6*(4), e18852.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno., et al. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: language-specific developmental changes in speech perception abilities. *1994 International Conference on Spoken Language Processing*, pp. 1695-1698. Yokohama: Japan.

Vihman, M. M. (2013). *Phonological development: The first two years*. John Wiley & Sons.

Yamada, R. A., & Tohkura, Y. I. (1992). The effects of experimental variables on the perception of American English/r/and/l/by Japanese listeners. *Perception & psychophysics*, *52*(4), 376-392.

# THE FORM AND FUNCTION OF CHIMPANZEE BUTTRESS DRUMMING

ELEUTERI VESTA[1,3], HENDERSON MATTHEW[1], SOLDATI ADRIAN[1,2], BADIHI GAL[1], ZUBERBÜHLER KLAUS[1,2,4], HOBAITER CATHERINE[1,4]

*Corresponding Author: vesta.eleuteri@univie.ac.at
[1]School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, Scotland, United Kingdom
[2]Department of Comparative Cognition, University of Neuchâtel, Neuchâtel, Switzerland
[3]Department of Behavioral and Cognitive Biology, University of Vienna, Vienna, Austria
[4]Budongo Conservation Field Station, Masindi, Uganda

## 1. Introduction

Human language is flexible. Speakers combine signal types (e.g., speech, gestures, facial expressions) to adjust the meaning of their messages. Signal combination is also present in non-human apes, but the functions of multi-signal combinations are still unclear (see Genty et al., 2014; Hobaiter et al., 2017; Wilke et al., 2017). Buttress drumming is a non-vocal acoustic signal often combined with pant-hoots, the species-typical long-distance vocalisation (Arcadi et al., 1998; Babiszewska et al., 2015). Chimpanzees produce drumming bouts by hitting the buttress roots of trees, generating low-frequency sounds that can be heard at over one kilometre (Arcadi et al., 1998). To understand why chimpanzees drum during pant-hoots, we investigated whether individual differences exist in the acoustic structure and in the timing within the pant-hoot of drumming bouts produced by adult male chimpanzees in different contexts and whether individual, contextual, and social factors impact their use of drumming.

## 2. Methods

Data were collected June-July 2019 and January-March 2020 in the Waibira chimpanzee community of the Budongo Forest Reserve, Uganda. During focal follows we noted the behavioural context of the focal individual and communication events produced while they were on the ground including a pant-hoot and/or drumming. We marked the identity of other individuals near the focal, representing the "party composition", and calculated the preferred social partners of each focal (following Babiszewska et al., 2015). Drumming events by any adult male were filmed and recorded *ad libitum.*

A 'Drumming bout' was defined as a series of beats produced by hitting the buttress roots of one tree with hands and/or feet (Arcadi & Wallauer, 2013). We coded five measures of acoustic structure for analysis (number of beats; bout duration; mean inter-beat duration; shortest inter-beat interval; longest inter-beat interval) and marked the start and end point of the drumming bout during the pant-hoot (Arcadi et al., 1998; Fedurek et al., 2016).

## 3. Results

We found individual differences in the acoustic structure of drumming bouts produced during traveling (n=105; DFA1: $p<0.001$) and in their timing during the pant-hoot. In contrast, we found no individual differences in the acoustic structure of drumming bouts produced by the same individuals during displays (n=36; DFA2: $p=0.589$). Pant-hoot drumming combinations were mainly present during traveling while virtually absent during displays. Chimpanzees drummed most frequently in the context of traveling and drummed less frequently as their party size increased. We found no effect of the social composition of the party on the use of drumming.

## 4. Discussion

Encoding of individual identity and increased use while traveling and when in smaller parties suggest that drumming may be incorporated in pant-hoots during traveling to facilitate chimpanzee fission-fusion spatial dynamics, specifically to join other individuals. In addition, absence of individual differences in display drumming suggests that chimpanzees can choose when to encode signaller identity in drumming "signatures", altering the function of this signal across social and behavioural contexts. By exploring the use of pant-hoot drumming this study helps elucidate the functions of multi-signal combinations in non-human apes, contributing to our understanding of the evolutionary origins of the flexibility central to human language.

## Acknowledgements

We thank the Uganda National Council for Science and Technology, the President's Office, the Uganda Wildlife Authority, and the National Forestry Authority for permission to conduct research, and the Royal Zoological Society of Scotland for their support of the field site. This research received funding from the European Union's 8th Framework Programme, Horizon 2020, under grant agreement no 802719.

## References

Arcadi, A. C., Robert, D., & Boesch, C. (1998). Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates*, *39*(4), 505–518. https://doi.org/10.1007/BF02557572

Arcadi, A. C., & Wallauer, W. (2013). They wallop like they gallop: Audiovisual analysis reveals the influence of gait on buttress drumming by wild chimpanzees (*Pan troglodytes*). *International Journal of Primatology*, *34*, 194–215. https://doi.org/DOI 10.1007/s10764-013-9656-1

Babiszewska, M., Schel, A. M., Wilke, C., & Slocombe, K. E. (2015). Social, contextual, and individual factors affecting the occurrence and acoustic structure of drumming bouts in wild chimpanzees (*Pan troglodytes*). *American Journal of Physical Anthropology*, *156*(1), 125–134. https://doi.org/10.1002/ajpa.22634

Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. *Scientific Reports*, *6*(1), 38226. https://doi.org/10.1038/srep38226

Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PLoS One*, *9*(1), e84738. https://doi.org/10.1371/journal.pone.0084738

Hobaiter, C., Byrne, R. W., & Zuberbühler, K. (2017). Wild chimpanzees' use of single and combined vocal and gestural signals. *Behavioral Ecology and Sociobiology*, *71*(6), 96. https://doi.org/10.1007/s00265-017-2325-1

Wilke, C., Kavanagh, E., Donnellan, E., Waller, B. M., Machanda, Z. P., & Slocombe, K. E. (2017). Production of and responses to unimodal and multimodal signals in wild chimpanzees, *Pan troglodytes*

*schweinfurthii*. *Animal Behaviour*, *123*, 305–316.
https://doi.org/10.1016/j.anbehav.2016.10.024

# LECTAL CONTACT AS A PATH TO LANGUAGE SPECIATION

T. MARK ELLISON*[1], LUISA MICELI[2], and JOHN MANSFIELD[3]

*Corresponding Author: t.m.ellison@uni-koeln.de
[1]Universität zu Köln, Cologne, Germany
[2]University of Western Australia, Perth, Australia
[3]University of Melbourne, Melbourne, Australia

Language speciation has been attributed to a loss of contact between parts of a speech community, or to contact with other languages (Thomason and Kaufman 1988). In this talk, we offer a psycholinguistically-grounded account for language speciation that does not presume a pre-existing fracturing of the linguistic community, or external contact.

Speciation begins with lectal variation being associated with an areal, social or situational identity. Three stages of differentiation follow. **Form differentiation** is an accelerated (compared to drift) loss of shared vocabulary in the lects due to bilectalism. **Structural stability** is maintained by the same bilectalism, even when lexical forms have differentiated beyond mutual intelligibility. If bilectalism breaks down, or is restricted in context, there is then rapid grammatical and semantic **structural divergence**.

This account explains why linguistic diversity in places with egalitarian multilingualism is apparent in lexical forms rather than structural features.

## 1. Lectal Variation

Lectal variation arises when social groups undergo fission, with an earlier shared lect dividing into distinct varieties. Early stages of social fission may involve social or geographic separation  (Trudgill 1986: 39); but distinct lects may also arise while the social group is in close contact (Labov 1963; Nettle & Dunbar 1997; Stanford 2009; Roberts 2010; Mansfield et al 2021). A shared Norman French variety on the isle of Jersey split when in the 1500s, forty families moved to nearby Sark. This social fission and the new local identities lead to lectal variation between Sark and Jersey dialects (Liddicoat 1994: 6).

## 2. Form Differentiation

Ellison & Miceli (2017) describe the anti-doppel bias (ADB) as a psycholinguistic process in which bilinguals prefer vocabulary not shared between their languages (i.e. non-doppels) where available. In contexts allowing both words, Dutch/English bilinguals prefer the English-only *picture* to the doppel *photo* in comparison to monolingual English speakers. The ADB seems likely to also apply in bilectals generally, including those who use multiple dialects. For example, Kapović (2005) notes that Croatian prescriptivists recommend *ponovo* rather than the common *ponovno* as it is (incorrectly) felt to particularly belong to the mutually-intelligible lect Serbian.

New lectal identities may develop slowly. Once, however, lects are associated with particular synonyms, the ADB can amplify differentiation, at least in the speech of bilectals. So long as many are bilectal, lexicon differentiation will increase. Mutual intelligibility will eventually break down.

### 3. Structural Stability

The next stage occurs while much of the population remain bilectal and the shared vocabulary has been substantially reduced (by Form Differentiation). The lects are no longer mutually intelligible and individuals identifying with one lect, learn the other as a foreign language.  The cognitive load using two lects,and monitoring for intrusive forms, facilitates cross-lectal structural priming, maintaining and stabilising structures shared between the lects. At this stage the lects essentially share one grammar but have distinct pools of lexical forms.

### 4. Structural Divergence

Where the social conditions promoting bilectalism weaken, or are lost completely (e.g. trade networks or marriage patterns change), contact between the lectal communities will be reduced or end. Without contact, there is no brake on differentiation of linguistic structures.

Because the lects by this stage have substantially distinct vocabularies, accidental similarities between lexical forms and semantics occur between different sememes. Analogical extension of constructions thus takes different paths in the two lects, leading to distinct morphological generalisations, and thus grammatical constructions peculiar to the lects, which are now different species.

### 5. Discussion

The above account of language speciation offers a number of advantages. It explains the shift from lectal differentiation to speciation in a known effect of human language processing, the ADB. It sees the structural uniformity but lexical form diversity seen in places like Vanuatu (François 2011) and Australia (Miceli 2019), as a natural consequence of egalitarian multilingualism never breaking down - the final stage of structural divergence is therefore never reached. Although we do not explore it here, it is possible to incorporate contact with distantly related or unrelated languages as part of this speciation model.

## References

Croft, W. 2000. *Explaining language change: an evolutionary approach*. London: Longman.

Ellison, T. M., & Miceli, L. (2017). Language monitoring in bilinguals as a mechanism for rapid lexical divergence. *Language*, 93(2), 255-287.

François, A. (2011). Social ecology and language history in the northern Vanuatu linkage: a tale of divergence and convergence. *Journal of Historical Linguistics* 1(2), 175-246.

Kapović, M. (2011). Language, ideology and politics in Croatia. *Slavia centralis*, 4(2), 45-56.

Labov, W. (1963). The social motivation of a sound change. *Word Journal of the International Linguistic Association* 19, 273–309.

Liddicoat, A. (1994). *A grammar of the Norman French of the Channel Islands*. Berlin: Mouton De Gruyter

Mansfield, J., Leslie-O'Neil, H. & Li, H. (2021). *Grammatical Variation and Dialect Divergence. New Ways of Analyzing Variation 49*. Austin, TX: University of Texas.

Miceli, L. (2019). *Unbalanced comparative patterns in historical linguistics: theoretical implications and explanatory mechanisms for their development.* [Doctoral dissertation, University of Western Australia].

Mufwene, S. (2007). Population movements and contacts in language evolution. *Journal of Language Contact*, 1(1), 63-92.

Mufwene, S. S. (2020). Genetic Creolistics as Part of Evolutionary Linguistics. *The Handbook of Historical Linguistics*, 2, 393-422.

Nettle, D., and Dunbar, R. I. M. Social Markers and the Evolution of Reciprocal Exchange. *Current Anthropology* 38(1): 93-99.

Roberts, G. (2010). An Experimental Study of Social Selection and Frequency of Interaction in Linguistic Diversity. In B. Galantucci and S. Garrod (Eds), *Experimental Semiotics: A New Approach for Studying the Emergence and the Evolution of Human Communication* (pp. 138–59). Amsterdam: John Benjamins.

Stanford, J. N. (2009). Eating the Food of Our Place': Sociolinguistic Loyalties in Multidialectal Sui Villages. *Language in Society* 38: 287–309.

Thomason, S. G. & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Trudgill, P. (1986). *Dialects in Contact*. Oxford: Blackwell.

# VOCAL ICONICITY IN NOMINAL CLASSIFICATION

NIKLAS ERBEN JOHANSSON*[1] and SANDRA CRONHAMN[1]

*Corresponding Author: niklas.erben_johansson@ling.lu.se
[1] Centre for Languages and Literature, Lund University, Lund, Sweden

While vocal iconicity in lexemes is increasingly well-studied (Blasi et al., 2016; Erben Johansson et al., 2020; Joo, 2020), its presence in more grammaticalized meaningful linguistic units are yet rather poorly understood. Aside from phonesthemes, vocal iconicity has been investigated in diminutives and augmentatives, but the results have been inconclusive (Körtvélyessy, 2011). Across languages, nominal classification involves fundamental semantic categories, such as sex (feminine/masculine), humanness (human/non-human), animacy (animate/inanimate), physical properties (size, shape) and functional properties (container, tool) (Aikhenvald, 2000). Thus, many of the categories found to be iconic in previous studies are also grammatically encoded in nominal classification systems.

The present study investigates the presence of vocal iconicity in nominal classification systems by using a genetically diverse dataset of languages, which was distributed evenly across the two main types of nominal classification systems (non-agreeing, such as more flexible *classifier* systems, and agreeing, such as more rigid *gender* or *noun class* systems). By collecting the nominal classification devices (NCDs) in 210 non-agreeing languages (126 language families) and 151 agreeing languages (123 language families), transcribing the NCDs using a coherent and comparable phonetic system, grouping them according to comparable semantic categories and analyzing them through Bayesian generalized linear models, we were able to assess whether certain types of sounds were overrepresented in certain class meanings.

The results revealed that the strongest overrepresentations of sounds were found in NCDs that pertained to shape and size (low, front, unrounded vowels in FLAT, high, back, rounded vowels in ROUND and high, front, unrounded vowels in SMALL) which aligned with previous cross-linguistic findings. However, the iconic effects were restricted to non-agreeing systems, and the non-agreeing systems included in the dataset contained more than three times as many nominal classes and almost twice as many segments as the agreeing systems.

These differences were attributed to more substantial phonetic erosion and semantic bleaching of agreeing NCDs through grammaticalization as nominal classification systems become increasingly formalized. This means that if a non-agreeing system changes into a more agreement-based system, it loses semantic transparency and gains formal predictability which dismantles one of the two key components of vocal iconicity and causes sound-meaning mappings to break down. While it is possible that these sound overrepresentations in non-agreeing systems exist solely because of inherited iconic effects present in the lexemes that the NCDs are derived from, iconicity has been shown to aid language acquisition (Imai & Kita, 2014; Massaro & Perlman, 2017; Nielsen & Dingemanse, 2021) which suggests that vocal iconicity could also be functional grammatical constructions. This would mean that the grammatical section of human language is affected by vocal iconicity and that the retrieval of noun semantics could be accelerated if primed with iconically congruent NCDs.

## References

Aikhenvald, A.Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

Erben Johansson, N., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*, *24*(2), 253–310. https://doi.org/10.1515/lingty-2020-2034.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Joo, I. (2019). Phonosemantic biases found in Leipzig-Jakarta lists of 66 languages. *Linguistic Typology*, *24*(1), 1–12. https://doi.org/10.1515/lingty-2019-0030.

Körtvélyessy, L. (2011). A cross-linguistic research into phonetic iconicity. *Lexis, 6*, 27–39. https://doi.org/10.4000/lexis.409

Massaro, D. W., & Perlman, M. (2017). Quantifying Iconicity's Contribution during Language Acquisition: Implications for Vocabulary Learning. *Frontiers in Communication, 2*. https://doi.org/10.3389/fcomm.2017.00004

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech*, *64*(1), 52–72. https://doi.org/10.1177/0023830920914339

# TONE AND LOANWORD ADAPTATION IN IXPANTEPEC MIXTEC: HORIZONTAL TRANSFER IN THE EVOLUTION OF LINGUISTIC DIVERSITY

DEANDRÉ A. ESPREE-CONAWAY [*1,2]

[*]Corresponding Author: despreec@uoregon.edu
[1]Department of Linguistic, University of Oregon, Eugene, Oregon,
United States of America
[2]Bloomfield Language Institute, Houston, Texas, United States of America

This poster describes phonological adaptation in Ixpantepec Mixtec [mks]—an Oto-Manguean language of southern Mexico. My main claim is that Ixpantepec Mixtec is incorporating Mexican Spanish loanwords into the language, adapting those loanwords into its phonology. This is a complex adaptive strategy for maintaining the vertically transmitted system while quickly and efficiently accepting new 'packages of information' useful for survival (of both the language and its speakers). The process is akin to 'conjugation' in horizontal gene transfer as it is a fitness- increasing strategy, found across the animal kingdom, although this specific configuration, predicated on combinatorial and phonotactic phonology is (human) language-specific. Special attention will be given to the placement of tones on loanwords. A corpus of over 200 loanwords is analyzed as part of this study.

## References

Chung, Y., & Ritchart, A. (2014a). Intonation patterns of interrogative constructions in Ixpantepec Nieves Tu'un Savi (Mixtec). *Journal of the Acoustical Society of America*, *135(4)*, 2291-2291.

Chung, Y., & Ritchart, A. (2014b). Question prosody in Ixpantepec Nieves Mixtec. (Oral presentation presented at SSILA 2014 (joint session with LSA), January 2014.

Crawford, C. (2008). An evolutionary account of loanword–induced sound change in Japanese. *University of Pennsylvania Working Paper in Linguistics*, *14(1)*, https://repository.upenn.edu/pwpl/vol14/iss1/5/ (6 December, 2021.)

Kager, R. (1999). *Optimality theory*. Cambridge: Cambridge University Press.

Kenstowicz, M., & Suchato, A. (2006). Issues in loanword adaptation: A case study from Thai. *Lingua*, *116*, 921-949.

Liang, B., & Du, Y. (2018). The functional neuroanatomy of lexical tone perception: An activation likelihood estimation meta-analysis. *Frontiers in Neuroscience*, *12(495)*, 1-17.

Monaghan, P., & Roberts, S.G. (2019). Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. *Cognition*, *186*, 147-158.

Peperkamp, S. (2005). A psycholinguistic theory of loanword adaptations. *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, *30(1)*, 341-352.

Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Malden: Blackwell.

Strachan, T., & Read, A.P. (2019). *Human molecular genetics (5th Edition)*. Boca Raton: CRC Press.

Tadmor, U. (2009). Loanwords in the world's languages: Findings and results. In Martin Haspelmath and Uri Tadmor (Eds.), *Loanwords in the world's languages: A comparative handbook* (pp. 55-75). Berlin: De Gruyter Mouton.

Truckenbrodt, H. (2016). Focus, intonation, and tonal height. In Caroline Féry and Shinichiro Ishihara (Eds.). *Oxford handbook of information structure* (pp. 463-482). Oxford: Oxford University Press.

Uchihara, H., & Mendoza Ruiz, J. (2021). Minimality, maximality and perfect prosodic word in Alcozauca Mixtec. *Natural languages & linguistic theory*, *1*, 1-51.

Uffmann, C. (2015). Loanword adaptation. In Patrick Honeybone and Joseph Salmons (Eds.), *The Oxford Handbook of Historical Phonology* (pp. 644-665). Oxford: Oxford University Press.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, *37*, 125-144.

# EXPLORING THE ROLE OF SPEECH IN THE TRANSMISSION OF AIRBORNE VIRUSES VIA A NOVEL METHOD FOR COMBINING AEROSOL AND PHONETIC ANALYSIS

CALEB EVERETT*[1], MARVA SEIFERT[2], CHANTAL DARQUENNE[2], RYAN TUMMINELLO[3] and JONATHAN SLADE[3]

*Corresponding Author: caleb@miami.edu
[1]Department of Anthropology, University of Miami, Coral Gables, USA
[2]School of Medicine, University of California San Diego, La Jolla, USA
[3]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, USA

Aerosol particle emissions from speech are thought to play a key role in airborne pathogen transmission, though relatively little is known about the mechanisms through which articulatory gestures produce aerosol particles. We describe a novel method that allows for the more precise isolation of the production of aerosol particles during speech. Some of the particles are smaller than those detected with previous approaches. Our approach allows for the isolation of aerosol bursts that are associated with specific articulatory gestures. Combined aerosol and phonetic analysis illustrates how this approach can shed light on the mechanisms through which aerosols are generated during speech. We briefly discuss ongoing work that has substantially refined this method. In relationship to language evolution, this work highlights the possibility that certain articulatory gestures could have "maladaptive" characteristics, in that they might generate inordinate numbers of aerosols. We stress that more experimental research is required to address this possibility.

## 1. Introduction

The role of speech in the production of airborne particles and associated pathogen transmission has recently been highlighted as epidemiological evidence implicates speaking in SARS-CoV-2 transmission. (Meselson 2020, van Doremalen et al. 2020). Previous studies have utilized several methods to examine how speech produces such particles, including laser-based detection, airflow testing and aerosol analysis. (Stadnytskyi et al. 2020, Asadi et al. 2020) The latter is essential as growing consensus exists that particles smaller than 10 μm in diameter, often referred to as aerosols, can transmit a variety of pathogens. (Fennelly 2020) The SARS-CoV-2 virus appears to remain viable in aerosols, which can remain airborne for hours, further increasing the odds of transmission. (Fennelly 2020) Research with an aerodynamic particle sizer (APS) has allowed scholars to show that some speakers are "super-emitters" of aerosols and that some sound types correlate with higher emission rates. (Asadi et al. 2019, 2020) However, APS instrumentation only samples once per second, limiting the extent to which it can isolate the role of specific articulatory gestures in producing aerosol particles. Other methods that have been used face related limitations. In English, words typically last 150-500 ms, while the most common syllable types last about 120-260 ms and individual sounds 60-150 ms. (Greenberg et al. 2003) Given the durations of such units in English and other languages, a more complete understanding of aerosol generation via speech could benefit from an approach with higher temporal resolution in aerosol measurements.

Here we utilize a method for detecting aerosols, one not previously applied to speech in the manner we illustrate, to address two lacunae in our current understanding of particles generated during talking: i) the size distribution of aerosols from 0.07 μm in aerodynamic diameter (i.e., at a size resolution of single viruses) to 10 μm and ii) the mechanics of the production of such fine aerosol particles on timescales relevant to individual sound types and utterances. Addressing these lacunae is critical to elucidating the full range of emission variations across speakers and sound categories and could potentially help refine the modeling of speech-based pathogen transmission. We outline a new approach that allows us to isolate with relative precision the moments at which aerosols emerge from speakers' mouths and the concentrations of aerosols speakers produce at various size bands. Phonetic analysis in tandem with aerosol detection reveals relationships between aerosol emissions and certain sound types not accessible by previous approaches. In this work we detail initial results based on three speakers, given pandemic-related restrictions to the number of participants we could test. This general method is being applied to dozens of participants in the coming months, though in current work participants breathe in particle-free air, in contrast to the results described here. The results discussed here simply serve as an illustration of the overall method, demonstrating how the heightened temporal and physical resolution of the approach offers potential gains to our understanding of the aerosol emissions associated with speech.

## 2. Results and discussion

For this study, three participants read various stimuli and breathed (nasally) at a natural rate into an aluminum funnel (opening diameter of 20 cm and length of 25 cm), connected to an electronic particle impactor that could measure the size and volume of aerosols produced ten times per second. Each individual's nose and mouth were within the funnel opening. The funnel was attached to a Dekati electrical low-pressure impactor (ELPI) via flexible conductive tubing with an inner diameter of 1.2 cm and a length of 25 cm. The short residence time in the tube and use of conductive tubing led to no distinguishable particle losses. This method detected aerosols at 10 Hz across 14 size bands, between 0.006 μm and

10 μm, though we focus only on those bands greater than 0.07 μm since they are relevant to viral transmission. Experiments were conducted in the absence of synthetic particle-free air to mimic a real-world setting. Background aerosol concentrations in the room air were measured before and after speaking or breathing. The mean background was then subtracted from the data measured while speaking or breathing. Although this method controlled for the background room aerosol by subtraction, in ongoing work we are using an approach with particle-free air. (Participants breathe in particle-free air and their exhaled air is mixed with particle free-air as well, rather than room air.) The key difference between the background-subtraction results is that the latter yield a much greater number of measured particles. Nevertheless, the results discussed here illustrate some of the gains of the new method, particularly the high physical and temporal resolution of the approach. Further, some of the results discussed here are also evident in our ongoing work employing particle-free air. For instance, across both approaches whispered sounds produce a higher number of particles than sounds produced with normal voicing at low amplitude.

In addition to the aerosol analysis, we also analyzed the utterances of the speakers acoustically. An InnoGear Condenser Professional Cardioid Microphone was placed about five cm to the right side of the funnel. (In our ongoing work, this is also combined with airflow analysis.) Speakers were given instructions and stimuli displayed on a desktop monitor placed immediately behind the funnel, so that they could read without moving their heads. Recordings were made onto a notebook computer at a sampling rate of 44.1 kHz. Relevant wav files were analyzed with PRAAT. A few dozen words were recorded for all speakers, in addition to the same paragraph. For this paper, speakers were asked to read at a "normal" volume and cross-speaker amplitude was not found to vary substantially. We did not have speakers read at intentionally varied volumes, as in previous work that uncovered an association between increased amplitude and increased aerosol emissions. (Asadi et al. 2019) Across all three speakers we observed greater aerosol production during speech than during normal breathing. Likely, greater differences in the concentrations of aerosols emitted between speaking and breathing occur when speaking at a higher amplitude or when singing, and minor amplitude differences across speakers may contribute to some of the inter-speaker variability we observed. (Asadi et al. 2019)

With the high temporal resolution of the method, we were able to assess aerosol production from specific sound types as highlighted in Fig. 1 for one speaker. In Fig. 1A we observe that, after mask removal, there are aerosol bursts immediately following each named letter of the recited alphabet. In this case the bursts of aerosols (diameter-weighted, *D*, μm cm$^{-3}$) occur primarily across two size bands of roughly 1-2 μm. Figure 1B illustrates how the temporal resolution of the method allows us to match aerosol bursts with individual syllables and even specific sound types when words are produced several seconds apart. Previous research suggests certain sound types, most notably the [i] high-front vowel, are associated with increases in aerosols, as is increased amplitude. (Asadi et al. 2020) Here we focused on consonants, including two consonant types that have previously been suggested to be relevant to particle emission. Given the temporal resolution of the method, we were able to directly observe aerosol bursts following some consonants in isolated words, including word-final consonants that have not previously been investigated. As illustrated in Figure 1B, at least in some words there was an increase in aerosols immediately following affricate consonants such as the last sounds in “catch” and “*h*” and the first sound in “*g*”. Figure 1C illustrates the variation in aerosol emissions across size bands during the articulation of one sentence previously used in research on the airflow produced during speech. (Abkarian et al. 2020) In this case there is an increase in aerosols shortly following the production of the [p$^{h}$], an aspirated bilabial plosive. This bilabial plosive has previously been shown to create an intense horizontal stream of air. (Abkarian et al. 2020) Some scholars have speculated that aspirated consonants like [p$^{h}$] may be relevant to airborne pathogen transmission via increased particle emission, but this is the first experimental evidence of any increase in aerosols associated with such aspiration. (Inouye 2003) We stress that it is very preliminary evidence, however, and must be replicated across many more participants with the new approach involving particle-free air. Initial results with the new approach are broadly consistent with those in Figure 1, suggesting that the association between aspiration and increased aerosols may be robust. Figure 1D further demonstrates the extreme concentration of aerosols that can be produced during some segments of speech for some speakers, in this case “shhh!”, a lengthened voiceless postalveolar fricative.

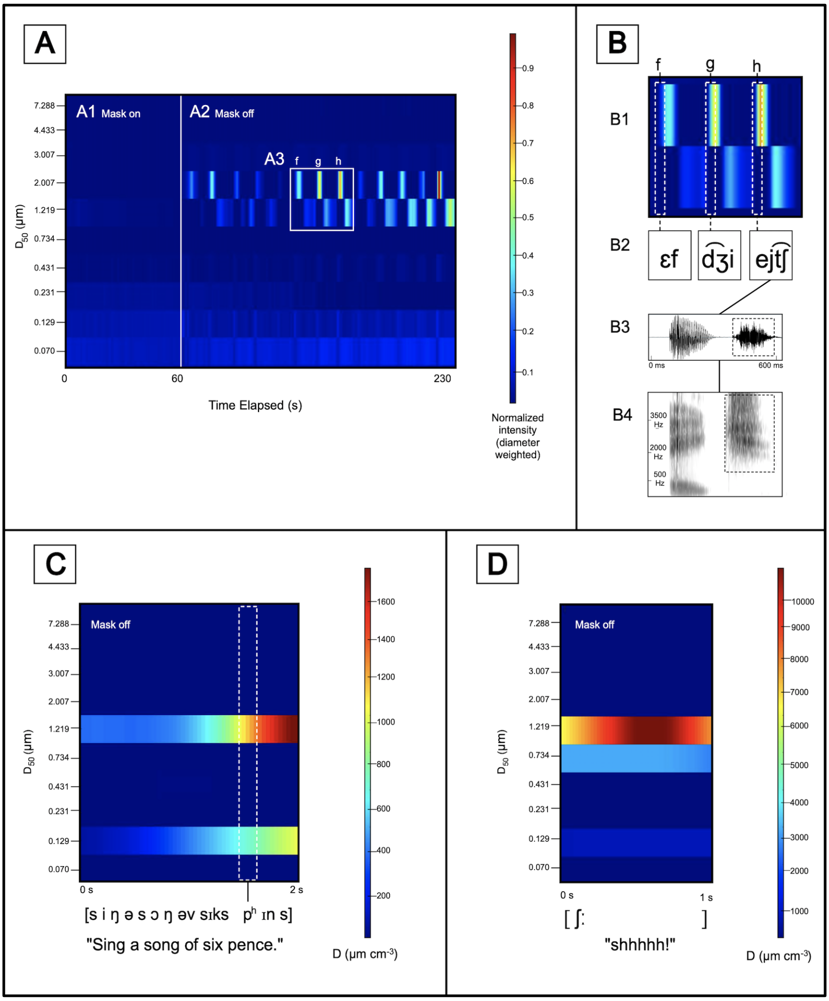


Figure 1. Temporal and physical resolution of method. Note that these values are much higher than the values obtained when speakers breathe in particle-free air, as demonstrated in our ongoing work. Panel A: Normalized heatmap of diameter-weighted aerosol particle concentration. A1. The speaker repeated the word "spar" at a normal volume, while wearing an N95 mask. A2. The speaker recited the alphabet at a normal volume, after removing his facemask. A3. Portion for *f*, *g*, and *h*, each said about eight seconds apart. Panel B: Section A3 enlarged, alongside phonetic details. B1. Aerosol heatmap. B2. Transcription of the three utterances with the International Phonetic Alphabet. B3. Waveform of [ejt͡ʃ]. Aperiodic sound waves associated with the voiceless affricate are highlighted. B4. Spectrogram of [ejt͡ʃ]. Aperiodic elements in high frequency range are highlighted. C. Aerosol visualization for one speaker's articulation of "Sing a song of six pence". D. Aerosols produced when the speaker was asked to say "shhh" (a voiceless postalveolar fricative), as though telling someone to be quiet. This was the most intense burst of aerosols for that speaker.

For more fluid segments of speech, the method still allows us to detect surges in aerosols associated with sounds or sequences of sounds though the aerosols do not occur in readily segmented bursts in such cases. It was observed that in fluid speech words with [st] sequences were weakly associated with increased emissions, an association that requires further exploration with a greater number of speakers and the particle-free approach. Such sequences, like the aforementioned affricates in "catch" and "shh" are of course voiceless. Previous work suggests that vocal cord vibration is also a key mechanism that dislodges and emits aerosol particles during speech (Asadi et al. 2019, 2020) but the high temporal resolution used here and in the laser-based detection in Stadnytskyi et al. (2020) suggests that at least some of the intense bursts of aerosols produced during speech are not produced in the larynx during vocal cord vibration. (In Stadnytskyi et al. (2020) it is observed that the voiceless interdental fricative in "healthy" emits many particles, though in a larger size range than that examined here.) It is possible that some sound types emit aerosols generated deeper in the respiratory tract, perhaps via the fluid-film burst mechanism in the terminal bronchioles. (Graham & Morawska 2009, Almstrand et al. 2010)

In short, the high resolution of this approach could be used to explore the detailed mechanisms through which aerosols are produced during speech, complementing other approaches. Ongoing refinement of this method could help to shed light not just on the temporal and physical dynamics of particle emission, but also on the mechanics through which aerosols are generated at the vocal cords and at other locations in the vocal tract during speech. Consistent with a growing literature using related methods that have more modest physical and temporal resolution, our results suggest that speaking does yield a high total number and volume of airborne particles that are potentially relevant to the transmission of some pathogens. Much work remains, however, to better understand how aerosols are produced during speech, along with the role of particularly articulatory gestures and associated sound types. In the context of language evolution, the preliminary results discussed here raise an interesting question: Do some articulatory gestures present a greater likelihood of intense aerosol bursts that can potentially transmit pathogens during an airborne pandemic? These results underscore this possibility, though we stress that much more experimental work is required, with a greater number of participants, to more fully understand whether some sound types have inaudible "maladaptive" features.

## References

Abkarian, M., S. Mendez, N. Xue, F. Yang, H. Stone. (2020). Speech can produce jet-like transport relevant to asymptomatic spreading of virus. *Proc. Natl. Acad. Sci. U. S. A.* 117, 25237-25245

Almstrand, A., B. Baker, E. Ljungstrom, P. Larsson, A. Bredberg, E. Mirgorodskaya and A-C Olin. (2010). Effect of airway opening on production of exhaled particles. *J. Appl. Physiol.* 108, 584–8.

Asadi, S., A. Wexler, C. Cappa, S. Barreda, N. Bouvier, W. Ristenpart. (2019). Aerosol emission and superemission during human speech increase with voice loudness. *Sci. Rep.* 9, 2348.

Asadi, S., A. Wexler, C.D. Cappa, S. Barreda, N. Bouvier, W. Ristenpart. (2020). Effect of voicing and articulation manner on aerosol particle emission during human speech. *PloS One* 15, e0227699.

Fennelly, K. (2020). Particle sizes of infectious aerosols: Implications for infection control. *Lancet Respir. Med.* 8, 914-924.

Graham, R., L. Morawska. (2009). The mechanism of breath aerosol formation. *Journal of Aerosol Medicine and Pulmonary Drug Delivery* 22, 229-237.

Greenberg, S., H. Carvey, L. Hitchcock, S. Chang. (2003).Temporal properties of spontaneous speech – A syllable-centric perspective. *Journal of Phonetics* 31, 465-485 (2003).

Inouye, S. (2003). SARS transmission: Language and droplet production. *The Lancet* 362, 170.

Järvinen, A., M. Aitomaa, A. Rostedt, J.Keskinen, J. Yli-Ojanperä. (2014). Calibration of the new electrical low pressure impactor (ELPI+). *J. Aerosol Sci.* 69, 150-159.

Meselson, M. (2020). Droplets and aerosols in the transmission of SARS-CoV-2. *The New England Journal of Medicine* 382, 2063.

Stadnytskyi, V., C. Bax, A. Bax, P. Anfinrud. (2020). The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proc. Natl. Acad. Sci. U. S. A.* 117, 11875-11877.

van Doremalen, N., D. Morris, M. Holbrook, A. Gamble, B. Williamson, A. Tamin, J. Harcourt, N. Thornburg, S. Gerber, J. Lloyd-Smith, E. de Wit, V. Munster. (2020). Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *The New England Journal of Medicine* 382, 1564-1567.

# SIGNAL EMBODIMENT ENHANCES GESTURE'S COMMUNICATION SUCCESS

Nicolas Fay*[1], Bradley Walker[1], T. Mark Ellison[2], Zachary Blundell[1], Naomi De Kleine[1],
Murray Garde[3], Casey J. Lister[1] and Susan Goldin-Meadow[4]

*Corresponding Author: nicolas.fay@gmail.com
[1]School of Psychological Science, University of Western Australia, Perth, Australia
[2]Collaborative Research Centre for Linguistic Prominence, University of Cologne, Cologne, Germany
[3]School of Culture, History and Language, Australian National University, Canberra, Australia
[4]Department of Psychology, University of Chicago, Chicago, USA

## Introduction

The origin of language of one of the oldest question in science, but vocal-first and gesture-first theories remain speculative due to a lack of direct evidence. Here we present empirical evidence to inform this debate.

Philosophers and early explorers recommended using '*the universal language of the hands'* to communicate with indigenous people (Quintilian 95CE; Cooperrider, 2019). Evidence for the universality of gesture was found in a recent paper that compared the communication success of gesture to (non-linguistic) vocalization cross-culturally (Australian and Ni-Vanuatu producers) and cross-experientially (sighted and severely vision-impaired Australian producers) (Fay et al., 2022). Like other referential communication studies (e.g., Ćwiek et al., 2021; Fay et al., 2013), communication success was measured by having a group of interpreters guess the meaning of each of the gesture/vocal signals produced (using a multiple choice format). Communication success was twice as high for the gestured signals than for vocal signals across the two experiments reported (61.17% versus 29.04%), and gesture's success was driven (in part) by its greater universality; the gestured signals produced by different participants were more similar in form than their vocal signals (i.e., they were more universal), and the degree of signal similarity was positively correlated with communication success.

Fay et al (2022) appealed to embodied theories of language and cognition to explain the success and universality of the gestured signals. Embodied theories highlight the importance of the body, and the body's interactions with the environment, to cognition (e.g., Clark, 1999; Hostetter & Alibali, 2008). Fay et al (2022) argued that whereas gesture lends itself to signal embodiment (e.g., communicating the word ‘drink’ via manually simulating raising a container to one’s mouth), the opportunity for signal embodiment is absent in the vocal modality.

Here, we re-analyzed the data collected by Fay et al (2022) to test an embodied account of the communication success of gestured signals. Producers communicated a large number of words (997 distinct words), an equal number of which were verbs, nouns and adjectives. In their analysis Fay et al (2022) did not distinguish between the different word categories. In the gesture modality, an embodied account predicts that verbs provide greater scope for signal embodiment than nouns, and nouns provide greater scope for embodiment that adjectives. If correct, communication success will be higher for verbs than nouns, and communication success will be higher for nouns than adjectives. In the vocal modality there is little scope signal embodiment, so no specific prediction is made with regard to the communication success of the different word categories. This is tested in the current study.

## Method

Re-analysis of the data collected by Fay et al (2022) and made openly available on the Open Science Framework: https://osf.io/36jpy/

## Results & Discussion

As predicted by an embodied account, in the gesture modality verbs were communicated more successfully than nouns and nouns were communicated more successfully than adjectives: 78%, 74%, 60% for Australian Producers, 64%, 55%, 46% for Ni-Vanuatu Producers and 60%, 48% and 38% for Severely Vision-Impaired Producers. No such pattern was evident in vocal modality: 33%, 31%, 34% for Australian Producers, 16%, 14%, 13% for Ni-Vanuatu Producers and 30%, 35% and 41% for Severely Vision-Impaired Producers. Our findings indicate that signal embodiment is important to gesture’s communication success. Similarities in the way people use gesture to embody meaning may also explain why gestured signals were found to be more universal than vocal signals.

## References

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences, 3*(9), 345–351. https://doi.org/10.1016/S1364-6613(99)01361-3

Cooperrider, K. (2019). Universals and diversity in gesture: Research past, present, and future. *Gesture*, *18*(2–3), 209–238. https://doi.org/10.1075/gest.19011.coo

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., Lupyan, G., Oh, G. E., Paul, J., Petrone, C., Ridouane, R., Reiter, S., Schümchen, N., Szalontai, Á., Ünal-Logacev, Ö., … Perlman, M. (2021). Novel vocalizations are understood across cultures. *Scientific Reports*, *11*(1), 10108. https://doi.org/10.1038/s41598-021-89445-4

Fay, N., Arbib, M., & Garrod, S. (2013). How to Bootstrap a Human Communication System. *Cognitive Science*, *37*(7), 1356–1367. https://doi.org/10.1111/cogs.12048

Fay, N., Walker, B., Ellison, T. M., Blundell, Z., De Kleine, N., Garde, M., Lister, C. J., & Goldin-Meadow, S. (2022). Gesture is the Primary Modality for Language Creation. *Proceedings of the Royal Society B: Biological Sciences*.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514.

# COMMUNICATION OR COGNITION? THE LOCUS OF RECURSION IN THE EVOLUTION OF HUMAN LANGUAGE

Andrew Feeney[1*]

[1]Department of Humanities, Northumbria University, Newcastle upon Tyne, United Kingdom
[*]andrew.feeney@northumbria.ac.uk

It is now more than 25 years since the publication of the ‘Minimalist Program’ (Chomsky: 1995) and this latest configuration of generative linguistics continues to exert considerable influence within the discipline and beyond. However, a major explanatory constraint against which such theories must now be evaluated has arisen in the form of the question of language evolution. This paper seeks to identify and resolve tensions in the notion of ‘recursion’ in relation to generative grammar and its interfaces with the conceptual-intentional (CI) and articulatory-perceptual (AP) or sensory-motor (SM) systems.

Writers working in different models of linguistics adhere to various accounts of how language evolved. In particular there is a distinction between those who propose a classic neo-Darwinian gradual narrative, and those, including Chomsky, who claim that the core aspect of language had a more or less instantaneous emergence in early *Homo sapiens*. I agree that positing an incremental evolution of such a mind internal capacity is deeply problematical. However, I argue that a saltationist account of the emergence of language, which seeks to identify a single key evolutionary step, is also flawed. The essential depiction of the evolutionary basis of a Minimalist theory of language appeared in Hauser, Chomsky and Fitch (2002) and was later clarified (Fitch, Hauser and Chomsky, 2005; Chomsky, 2005). In this account the uniquely human aspect of language was defined as the narrow recursive operation ‘merge’ by which lexical items were combined in hierarchical derivations.

The use of ‘recursion’ is not always well-defined and has resulted in considerable debate in the literature (see e.g. Jackendoff & Pinker, 2005; and papers in Lowenthal & Lefebvre, 2014). Nevertheless, two defining qualities can be identified. Firstly, there is the condition that the output of an earlier stage must constitute the input to a subsequent one, as in Fibonacci series. Secondly, often highlighted in relation to language, is the capacity for one element of type X to be embedded in an element of the same type. However, while hierarchical recursion certainly *appears* to be inherent in linguistic ‘structure’, it is also clearly evident in other aspects of modern human cognition, most clearly in the capacity by which I am able to have a thought about your belief about someone else’s belief, and so on. While primates, like all higher animals, have a rich conceptual system and a basic ability to reason based on observable cause and effect relations, there is undoubtedly, to a greater or lesser degree, a ‘mental gap’ between humans and other hominins and non-humans, often described as ‘intentionality’ or ‘theory of mind’ (e.g. Baron-Cohen, 1995). Furthermore, within cognitive science there are many compatible perspectives on human thought (CI system) that envisage individuals having a conceptual array and a system for structuring concepts independent of (externalised i.e. articulated) language, including ‘language of thought’ (Fodor, 1975, 2008) and similar hypotheses (e.g. Carey, 2011; Harnard, 2010; Wyn et al, 2009).

The question then arises of the locus of recursion and its evolutionary origin in relation to human language. While it has been suggested that this recursive property in cognition may be the source for the apparent recursion in language (‘an optimal solution to expressing recursive thought’, Kinsella, 2009: 152), there are no empirical or theoretical grounds for positing the existence of an autonomous level of linguistic syntactic structure in any sense. If there is a wholly internal system for the recursive hierarchical structuring of conceptual material (i.e. into thoughts), that constitutes the human CI system, then all that is additionally required is a separate system for the communication of those thoughts to others, the AP/SM system. This, I argue here, is effected by means of a wholly external system of semiotic representation as conceived in the Representational Hypothesis (RH) (e.g. Burton-Roberts, 2011). A central tenet of the RH is that there are essentially only two elements – sound (sign) and thought (signified) – in what is traditionally understood as language. In interpreting the word ‘cat’ the phonological string /kæt/ leads directly to a mental representation of the concept [CAT]; there is no need for intermediate categories such as a noun, a ‘syntactic’ object, nor any reason why such a level of representation would have evolved (either gradually or abruptly). The same is true of composite multi-conceptual entities including full propositions. Devitt attributes this thesis that ‘representations are not to be multiplied beyond necessity (2006: 51) to Pylyshyn and, borrowing from Occam, terms it ‘Pylyshyn’s Razor’. There is simply no evolutionary rationale for positing a syntactic architecture, neither cognitive nor communicative; and certainly not (contra widespread assumption) for the purpose of disambiguation, which is entirely mental.

Pursuit of the origins of language in the style of the MP, the quest for the source of *linguistic* recursion, can only hinder further insight into language evolution.

## References

Baron-Cohen, Simon. (1995) *Mindblindness: an essay on autism and theory of mind*. Cambridge, Mass.: MIT Press

Burton-Roberts, N. (2011) On the grounding of syntax and the role of phonology in human cognition. *Lingua*. 121, 2089-2102.

Carey, S. (2011) Precis of The Origin of Concepts. *Behavioral and Brain Sciences*.

Chomsky, Noam. (1995) *The Minimalist Program*. Cambridge, MA.: MIT Press

Chomsky, N. (2010) Some simple evo devo theses: How true might they be for language? *In:* Larson, R. K., Deprez, V. & Yamakido, H. (eds.) *The evolution of human language: Biolinguistic perspectives*. Cambridge: Cambridge University Press.

Devitt, M. (2006) *Ignorance of language*. Oxford: Oxford University Press

Fitch, W. T., Hauser, M. D. & Chomsky, N. (2005) The evolution of the language faculty: Clarifications and implications. *Cognition,* 97**,** 179-210.

Fodor, J. (1975) *The language of thought*. New York: Crowell

Fodor, J. (2008) *LOT 2: The language of thought revisited*. Oxford: Oxford University Press

Harnad, S. (2010) Eliminating the "concept" concept. *Behavioral and Brain Sciences*. 33, 213-+

Hauser, M. D., Chomsky, N. & W. Fitch, T. (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science*. 298, 1569-1579.

Jackendoff, R. & Pinker, S. (2005) The nature of the language faculty and its implications for evolution of language. Cognition. 97, 211-225

Kinsella, A. (2009) *Language Evolution and Syntactic Theory*. Cambridge: Cambridge University Press

Lowenthal, F. & Lefebvre, L. (eds) (2014) *Language and Recursion*. London: Springer

Wynn, T., Coolidge, F. & Bright, M. (2009) Hohlenstein-Stadel and the Evolution of Human Conceptual Thought. *Cambridge Archaeological Journal*. 19, 73-83

# A GENERAL SOLUTION TO THE COMBINED EFFECTS OF SELECTION AND BIASED MUTATION IN CULTURAL EVOLUTION

Vanessa Ferdinand

Melbourne School of Psychological Sciences, University of Melbourne, Australia
Santa Fe Institute, New Mexico, USA

The evolution of culture is shaped by two primary forces of directional change: *biased transmission*, which is similar to natural selection, and *biased reconstruction*, which is similar to non-random recurrent mutation. The relative strength of these two forces has been the subject of much debate and formal models to date provide conflicting results with no definitive resolution. In this paper, I derive an analytical solution to the combined effects of selection and biased mutation when each are formalized as stochastic processes (e.g. Griffiths & Kalish, 2007). An important observation that immediately appears in the solution is that the temporal order in which these forces operate causally affects the evolutionary dynamics of the system. This order effect explains conflicting results in the literature to date: models in which selection appears to be stronger allow selection to operate last in the generational structure (e.g. Henrich & Boyd, 2002) and models in which biased mutation appears to be stronger allow mutation to operate last (e.g. Griffiths, Kalish, & Lewandowsky, 2008). In cultural evolution, various combinations of biasing forces are possible and these combinations are likely to vary across empirical domains. The solutions in this paper offer a shortcut to calculating the combined effects of any arbitrary number and ordering of selection and biased mutation processes, which should facilitate a more complete characterization of the space of dynamics that cultural evolution is capable of and improve our understanding of the relative effects of selection and biased mutation processes in culturally evolving systems.

## 1. Introduction

The evolution of culture is shaped by two primary forces of directional change. The first type is selective in nature; it determines which kinds of cultural entities will be copied by people (winning these variants fecundity and “offspring”) or ignored (effectively pronouncing these variants “dead”). A variety of selective forces operate on culture, such as natural selection, social learning strategies, and biased transmission processes (Boyd & Richerson, 1985). The second type is mutational in nature; it transforms cultural entities from one type to another while these entities are being copied and is capable of injecting novel variation into culture, often in biased ways. A variety of mutation-like processes operate on culture and fall under the umbrella of biased reconstruction processes (Sperber, 1996). Both of these forces are capable of creating directional changes that increase the frequency of certain cultural variants relative to others.

There has been a great deal of debate over the relative importance of each of these forces in shaping culturally evolving systems (e.g. Acerbi & Mesoudi, 2015). Although much of the discussion focusses on the fidelity of the cultural copying process, such as high vs low rates of random mutation, researchers are also interested in pitting biased mutation against selective forces and observing which force is more causally potent. This latter endeavour has produced several idiosyncratic models with conflicting results, leading Acerbi and Mesoudi (2015) to suggest the debate be moved to the empirical realm.

However, there is still conceptual ground to be gained in the modelling realm. What is lacking is a sufficiently general framework that allows a wide range of fitness functions to be combined with a wide range of biased mutation functions and straightforward analytical solutions to each of their combinations. In this paper, I adopt a general stochastic process modeling framework for cultural evolution and derive an analytical solution for the combined effects of any arbitrary fitness values and any arbitrary biased recurrent mutation process.

## 2. Modelling framework

This section proposes a stochastic process model (e.g. Lawler, 2018) for biased cultural mutation. Stochastic process modelling is commonly used in complex systems science and was first applied to cultural evolution in the domain of language evolution (Niyogi & Berwick, 1997; Nowak, Komarova, & Niyogi, 2001, 2002) and expanded with the development of the Bayesian iterated learning model (Griffiths & Kalish, 2007). The main representational device in this framework is a stochastic matrix called a *transition matrix* which contains the probability, $p_{ij}$, that a variable in state $v_i$ transitions to state $v_j$:

$$\mathbf{M} = \begin{array}{c} \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & \dots & v_n \end{array} \\ \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,j} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i,1} & p_{i,2} & \cdots & p_{i,j} \end{pmatrix} \end{array} \tag{1}$$

For the case of biased mutation, $p_{ij}$ is be the probability that a cultural variant of type $v_i$ mutates into type $v_j$ and we will refer to the transition matrix as a *mutation matrix*, $\mathbf{M}$. Transition matrices are also used in the field of mathematical population genetics to represent non-random recurrent mutation processes (e.g. Ewens, 2004, p. 174), but these are generally only 2-allele models with 2 × 2 matrices. Cultural entities, however, typically exhibit several variants. For example, if you go to the hardware store to decide which cultural variant of "door handle" you'd like your kitchen cabinets to have, there won't be only two options, but dozens of variants to choose from. Equation 1 accommodates this by generalizing to an arbitrarily large number of variants, $V = \{v_1, v_2, ...v_n\}$.

Matrices are useful organizational devices for visualizing different patterns of mutation flow. The off-diagonal values on the $i$th row show $v_i$'s losses under the mutation process and the off-diagonal values on the $j$th column show $v_j$'s gains. The probabilities on each row sum to one, indicating that all of the mutational events are accounted for, but the columns may or may not sum to one. When the column sums do not all sum to one, the mutation process is *biased*, with certain variants receiving higher gains under the mutation process than others.

Representing biased mutation by a stochastic matrix is also convenient because it puts the tools of linear algebra at our disposal for calculating things such as the evolutionary trajectories of populations and their equilibrium frequencies. To make use of these tools, we represent the population as a stochastic vector, $\vec{q_t}$, where $\vec{q_t}(v_i)$ is the relative frequency of $v_i$ in the population at time $t$. For example, $\vec{q_t} = (.3, .5, .2)$ means that $v_1$ constitutes 30% of the population at time $t$ whereas $v_2$ constitutes 50% and $v_3$ constitutes 20%. We will refer to $\vec{q_t}$ as the *population vector*. The expected frequency of each variant in two successive generations is $\vec{q}_{t+1} = \vec{q_t}\mathbf{M}$ and for any arbitrary number of generations into the future, $g$, it is $\vec{q}_{t+g} = \vec{q_t}\mathbf{M}^g$. The equilibrium frequencies are found by solving for the stationary distribution(s) of $\mathbf{M}$, as in $\pi = \pi\mathbf{M}$, or by solving $\pi = \vec{q_t}\mathbf{M}^\infty$, using a sufficiently large numeric stand-in for $\infty$. Example solutions are shown in Figure 1 (leftmost panel) for the example mutation matrix

$$\mathbf{M} = \begin{pmatrix} .8 & .1 & .1 \\ .2 & .6 & .2 \\ .3 & .3 & .4 \end{pmatrix} \tag{2}$$

where each black line represents the expected frequency of $v_i$ over the course of 50 generations and they converge to the unique stationary distribution of $\mathbf{M}$, which is $\pi = (.55, .27, .18)$. Here we see that $\mathbf{M}$ is a biased process that causes $v_1$ to be the most plentiful variant, constituting 55% of the population at equilibrium.

The stochastic process model of cultural mutation described here is more general than the Bayesian iterated learning model because it does not require cultural mutation to be implemented by Bayesian rational learners, and it is more restricted than the Evolutionary Causal Matrix developed by Claidière, Scott-Phillips, and Sperber (2014) because 1) the entries in the matrix are probabilities and 2) it is meant to model a non-random recurrent mutation *process*, rather than model cultural attraction theory more broadly.

## 3. Combined effects solution

One approach to deriving an analytical solution to the combined effects of selection and biased mutation entails finding a way to incorporate fitness values into the mutation matrix, creating a new combined effect matrix, $\mathbf{C}$, and then applying our linear algebra solutions to **C**. In this section, I show how this can be done for a general model of natural selection, the replicator dynamics (Nowak, 2006).[1] The discrete-time replicator equation can be written as $q_{t+1}(v_i) \propto q_t(v_i)f(v_i)$, where $f(v_i)$ is the fitness of the $i$th variant. Fitness values can be represented in several ways: as each type's expected number of children, e.g. $f(v_1, v_2) = (10, 5)$, as each type's fitness relative to one a reference type, e.g. $f(v_1, v_2) = (1, \frac{1}{2})$, or as each type's fitness relative to the population's total fitness, e.g. $f(v_1, v_2) = (\frac{2}{3}, \frac{1}{3})$, but the choice of representation does not affect the solution for the relative frequency $q_{t+1}(v_i)$. We will work with the latter representation where fitness is a stochastic vector and refer to it as the *fitness vector*. Figure 1 (rightmost panel) shows the dynamics for example fitness values $f(v_i) = (.5, .4, .1)$. Here we see that the variant with the highest relative fitness, $v_1$, goes to fixation, meaning it takes over the population and drives all other variants to extinction.

The combined effects matrix, **C**, can be obtained by converting the fitness vector into a diagonal matrix, $F$, where

$$\mathbf{F} = diag(f_1, ..., f_n) := \begin{bmatrix} f_1 & & \\ & \ddots & \\ & & f_n \end{bmatrix} \tag{3}$$

and then multiplying **F** by the mutation matrix, **M**. It is important to note that matrix multiplication is not commutative, meaning $\mathbf{F} \times \mathbf{M} \neq \mathbf{M} \times \mathbf{F}$. Therefore, there are clearly two solutions to **C**: $\mathbf{C}_{fm} = \mathbf{F} \times \mathbf{M}$, where selection acts before mutation in the generational structure of the cultural system, and $\mathbf{C}_{mf} = \mathbf{M} \times \mathbf{F}$, where mutation acts before selection.

When selection occurs first, $\mathbf{F} \times \mathbf{M}$ implements element-wise multiplication of the fitness vector by each *column* in **M** and can be understood as incorporating the fitness vector into the *gains* or *input* values of the mutation matrix, allowing selection to act before the mutation process occurs:

$$\mathbf{C}_{fm} = \begin{bmatrix} f_1 p_{1,1} & f_1 p_{1,2} & \cdots & f_1 p_{1,j} \\ f_2 p_{2,1} & f_2 p_{2,2} & \cdots & f_2 p_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ f_n p_{i,1} & f_n p_{i,2} & \cdots & f_n p_{i,j} \end{bmatrix} \tag{4}$$

When mutation occurs first, $\mathbf{M} \times \mathbf{F}$ implements element-wise multiplication of the fitness vector by each *row* in **M** and can be understood as incorporating

---

[1] Even though it contains the word "replicator", the replicator dynamics equation itself neither implies nor requires that the variants in question be high-fidelity replicators. The equation simply assigns fitness values to variants based on their type at the moment the selection process acts and can be combined with any model of mutation.

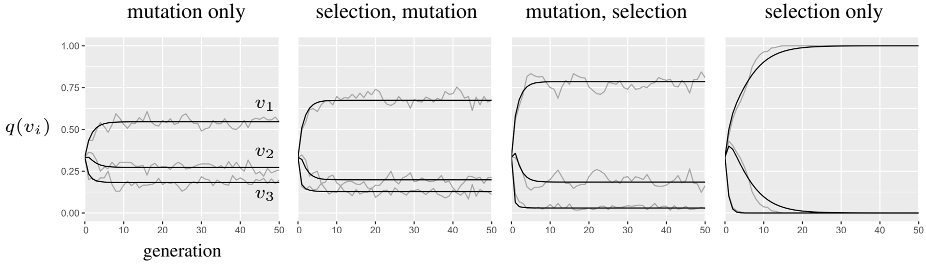


Figure 1. Evolutionary dynamics of four example combinations of selection and biased mutation, using fitness values $f(v_i) = (.5, .4, .1)$, mutation process $\mathbf{M}$ from Equation 2, and initial population $q(v_i) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Each panel (from left to right) shows the result when: mutation acts alone, selection occurs before mutation in the generational structure, mutation occurs before selection, and selection acts alone. Black lines show the expected frequency of each variant in the population at the end of each generation: $v_1$ (top), $v_2$ (middle), $v_3$ (bottom). Grey lines show the result of a simulated trajectory for a finite population size of 300 variants. The expected trajectories converge to the stationary frequency of each variant, $\pi(v_i)$.

the fitness vector into the *losses* or *output* values of the mutation matrix, allowing selection to act after the mutation process occurs:

$$\mathbf{C}_{mf} = \begin{bmatrix} f_1 p_{1,1} & f_2 p_{1,2} & \cdots & f_n p_{1,j} \\ f_1 p_{2,1} & f_2 p_{2,2} & \cdots & f_n p_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ f_1 p_{i,1} & f_2 p_{i,2} & \cdots & f_n p_{i,j} \end{bmatrix} \tag{5}$$

Figure 1 shows the results for $\mathbf{C}_{fm}$ (second panel) and $\mathbf{C}_{mf}$ (third panel) for identical **F** and **M**. Here we see that the dynamics of these two processes differ. Process $\mathbf{C}_{fm}$ more closely resembles the dynamics of mutation acting on its own, whereas $\mathbf{C}_{mf}$ seems to be more influenced by selection. The analytical solutions $\pi\mathbf{C}_{fm} = (.67, .20, .13)$ and $\pi\mathbf{C}_{mf} = (.79, .18, .03)$ confirm that these are, indeed, identifiably different stochastic processes.

## 4. Discussion of order effects

In the solution above, we saw that order matters because matrix multiplication is not commutative. To understand why order matters, let's consider the following thought experiment. Imagine that you and I are standing in front of a ball of clay. What we're going to do is take turns punching the ball of clay: first I throw a punch, then you throw a punch, and we repeat this a few more times. Now let's both inspect the clay and ask, whose handprint can we see most clearly? The one of us whose punch is the strongest? Perhaps. Or what about the one of us who punched the clay most recently? If we introduce a consistent bias into our measurement system by only inspecting the clay after I punch it, then we are

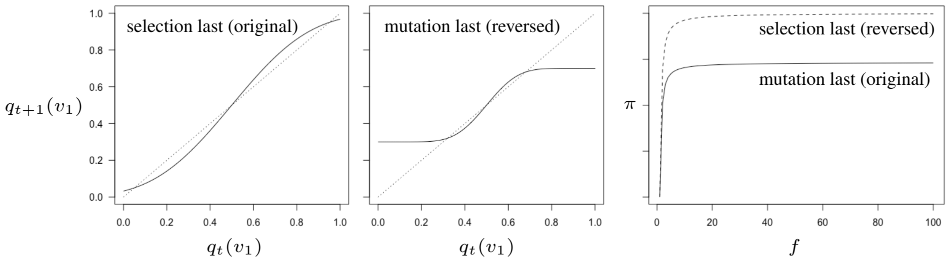


Figure 2. Replications of Henrich and Boyd (2002) (left, middle) and Griffiths et al. (2008) (right) alongside results that reverse the order of selection and mutation in their original models.

probably going to walk away thinking I packed the stronger punch. But if we only inspect it after you punch it, then we'll make the opposite inference.

Many formal models of evolution assume discrete generations and those that deal with multi-process evolutionary systems normally organize the processes into a set linear order within the generational structure. Furthermore, population frequencies are typically measured at the end of the generation, rather than after each force acts. This particular set of modeling choices is common, but it is also a recipe for over-representing the effects of each generation's final force.

Two examples of this ordering effect bias from the cultural evolution literature are found in models by Henrich and Boyd (2002) and Griffiths et al. (2008). Both of these papers address the relative potency of selection vs mutation in cultural evolution, but their models yield opposite results. Henrich and Boyd (2002) combine a model of conformity copying (a type of selection) with high mutation rates and find that the resulting evolutionary dynamics are barely affected by mutation. Figure 4 (left) replicates their result for a mutation rate of 30% and a population size of 20 (see their Figure 5). The overall shape of this curve resembles the typical s-shaped curve of conformity copying, which would have stable equilibria at coordinates $q(v_1) = 1$ and 0. When combined with high mutation rates, the stable equilibria remain near $q(v_1) = 1$ and 0 and do not display the 30% mutational load that would be expected from a 30% mutation rate. On the basis of this model, the authors conclude that error-prone cultural evolution is "corrected" by conformity biases. However, in this model, mutation operates first in the generational structure, conformity copying operates last, and frequencies are measured at the end of each generation. Figure 4 (middle) shows the results when the order of the selection and mutation processes are reversed. Here, the effects of mutation are much more pronounced and we observe the expected mutational load of 30%, pushing the equilibria to $q(v_1) = 0.7$ and 0.3. This new perspective suggests a win for mutation over conformity copying.

Griffiths et al. (2008) combine a model of cultural mutation, resulting from a Bayesian learner with a prior bias favoring one cultural variant ($v_1$) over another

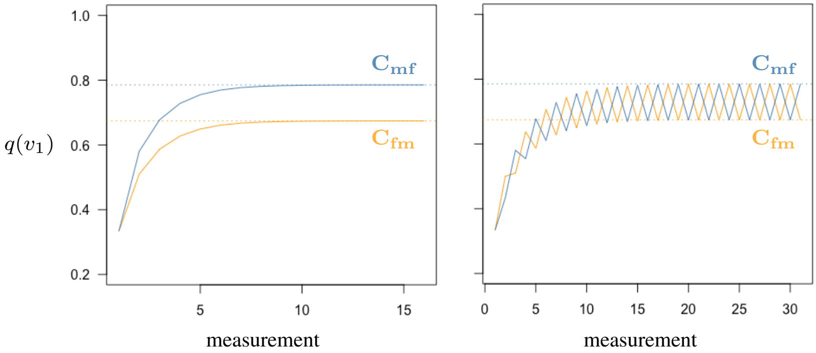


Figure 3. Replot of the trajectories defined by process $\mathbf{C}_{\mathbf{mf}}$ (blue) and $\mathbf{C}_{\mathbf{fm}}$ (orange) showing the frequency of $v_1$ over 15 generations when measured at the end of each generation (left) and when measured twice per generation after each force acts (right). Both processes begin with the same uniform initial population composition, $\vec{q}_{t=0} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

($v_2$), with a selection pressure causing differential fitness among the variants. This model also assumes a fixed generational structure, but one where selection operates first and mutation operates last. Figure 4 (right, solid line) replicates their result for a $20\%$ prior bias on $v_1$ and a $10\%$ noise rate (corresponding to parameters $\pi = 0.2$ and $\epsilon = 0.1$ in their notation, see their Figure 5). Here, the equilibrium frequency of $v_1$ ($x$-axis) goes up as the fitness of $v_1$ increases ($y$-axis), but plateaus around $78\%$ due to a mutational load of about $22\%$. This result was interpreted as a win for mutation over selection, because even infinitely strong selection is incapable of pushing $v_1$ to fixation. The dashed line, however, reverses the order of these two processes. In this case, $v_1$ goes to fixation as fitness goes to infinity and this scenario could be interpreted as a win for selection over mutation.

How should we deal with these ordering effects when evaluating the relative effects of selection and biased mutation in cultural evolution? One solution is to simply measure the population after every process acts. Referring back to Figure 1, the data in these plots show the population frequencies at the end of each generation only. Figure 4 replots the evolutionary trajectories of $v_1$ for process $\mathbf{C}_{\mathbf{fm}}$ and $\mathbf{C}_{\mathbf{mf}}$, where the system is measured once at the end of each generation (left) and once after each force acts (right). The right panel clearly shows that each force affects the frequency of $v_1$ when it acts and that the system reaches an oscillating attractor where the two forces kick the same population frequencies back and forth between the equilibrium frequency solutions for $\mathbf{C}_{\mathbf{fm}}$ and $\mathbf{C}_{\mathbf{mf}}$.

A second solution is to give a concrete or theoretically-justified interpretation to the ordering of selection and mutation processes in our models. This is more challenging as it requires some understanding about the “life-cycle” of a cultural variant, including when specific selection and mutation forces typically act upon it. For example, biases in attention, memory, and the pragmatic usage of language may occur in a set order that is causally relevant to the evolution of language. Future research into the relative importance of selection and biased mutation in cultural evolution should either control for order effects or explicitly address the causal role they play in determining population frequencies over time.

## Acknowledgements

This work was completed with support from the University of Melbourne's Research Fellowship in Computational Cognitive Science, under the supervision of Andrew Perfors and Yoshihisa Kashima, and the Omidyar Fellowship at the Santa Fe Institute. Thanks to three anonymous reviewers for their helpful comments.

## References

Acerbi, A., & Mesoudi, A. (2015). If we are all cultural darwinians what's the fuss about? clarifying recent disagreements in the field of cultural evolution. *Biology & philosophy*, *30*(4), 481–503.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process.* University of Chicago Press.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(20130368).

Ewens, W. J. (2004). *Mathematical population genetics: theoretical introduction* (Vol. 1). Springer.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive science*, *31*(3), 441–480.

Griffiths, T. L., Kalish, M. L., & Lewandowsky, S. (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3503–3514.

Henrich, J., & Boyd, R. (2002). On modeling cognition and culture. *Journal of Cognition and Culture*, *2*(2), 87–112.

Lawler, G. F. (2018). *Introduction to stochastic processes.* Chapman and Hall.

Niyogi, P., & Berwick, R. C. (1997). Evolutionary consequences of language learning. *Linguistics and Philosophy*, *20*(6), 697–719.

Nowak, M. A. (2006). *Evolutionary dynamics: exploring the equations of life.* Harvard university press.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science*, *291*(5501), 114–118.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, *417*(6889), 611–617.

Sperber, D. (1996). Explaining culture: A naturalistic approach. *Cambridge, MA: Cambridge*, *1101*.

# BABBLING BEHAVIOR: SHARED VOCAL ONTOGENETIC FEATURES IN BAT PUPS AND HUMAN INFANTS

AHANA A. FERNANDEZ*[1], LARA S. BURCHARDT[1], MARTINA NAGY[1] AND MIRJAM KNÖRNSCHILD[1,2,3]

*Corresponding Author: fernandez.aurora.ahana@gmail.com

[1]Natural History Museum – Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

[2]Institute of Biology, Freie Universität Berlin, Berlin, Germany

[3]Smithsonian Tropical Research Institute, Panama City, Panama

Speech is the vocal motor output aspect of language and requires precise control over a complex muscle system (e.g. laryngeal and orofacial muscles) steered by neuronal mechanisms. During ontogeny, every child is challenged with the acquisition of speech sounds. The first utterances resembling speech sounds are observed during babbling, a distinctive vocal behavior in human infants (Oller, 2000). Babbling is a production milestone in infant speech development because it enables infants to practice speech sounds by gaining control over their speech articulators (Oller, 2000; ter Haar et al., 2021; Vihman, 2014).

With our study, we introduce a new and promising mammalian candidate for comparative biolinguistic research on vocal ontogenetic processes, the greater sac-winged bat *Saccopteryx bilineata.* The bat *S. bilineata* is a vocal production learner (Knörnschild, Nagy, Metz, Mayer, & von Helversen, 2010) and exhibits a conspicuous vocal practice phase during ontogeny (Fernandez, Burchardt, Nagy, & Knörnschild, 2021; Knörnschild, Behr, & von Helversen, 2006). This pup vocal practice is strongly reminiscent of infant babbling: it is organized in bouts composed of multisyllabic repetitive sequences that are interspersed by short silent intervals. Babbling bouts can last up to 43 minutes and throughout the babbling phase (i.e. the period between first and last day of babbling), pups spend about 30% of their active time with babbling. During babbling pups learn to produce song by imitating adult tutors; pups' renditions of adult song are often found in babbling bouts (Fernandez et al., 2021; Knörnschild et al., 2006; Knörnschild et al., 2010).

In human infants, irrespective of the language to be learned, babbling is characterized by several common features, e.g. repetitiveness, rhythmicity, early onset in infancy (Oller, 2000; Vihman, 2014). Our study investigated whether babbling behavior in *S. bilineata*, pups is characterized by the same features that define infant babbling. The aim was to provide the first formal comparison of babbling features across vocal learning mammals.

For this study, we investigated the entire vocal ontogeny of 20 pups from two wild populations in Costa Rica and Panama. Pup babbling was organized in three hierarchical levels: the syllables (level 1) were concatenated to sequences called syllable trains (level 2), which – interspaced by silent intervals – constituted the babbling bouts (level 3). On the syllable level, we investigated the following features: age at babbling onset (A), babbling bout composition (B), syllable type acquisition (C), syllable type emergence during the babbling phase (D), reduplication (E), social context and function of babbling (G) and universality (H). We analyzed at least one babbling bout per week and pup (N=216 babbling bouts) for investigating all parameters, except for reduplication (E). To investigate (E) we classified 55´056 syllables of 10 pups into syllable types and calculated a correlation matrix showing the observed repetition rates of the syllable types present in babbling. Rhythmicity (F) was investigated on the syllable train level (N=30 trains, 712 syllables) by measuring syllable inter-onset intervals and further assessed with a normalized pairwise variability index. For (A-D) and (H) we visually classified syllable types in babbling bouts based on their high spectro-temporal similarity with the adult vocal repertoire (which is entirely delineated). Our visual classification was statistically verified by measuring a subset of our data and performing discriminant function analyses (acoustic analysis: Avisoft SasLab Pro, DFA: SPSS).

Our results revealed that the features that define human infant babbling are also characteristic of pup babbling: (A) babbling onset occurred early, at one third of the entire vocal ontogeny, (B) babbling bouts were composed of adult-like syllable types (comparable to infant canonical syllables) and undifferentiated proto-syllables (comparable to infant speech precursors). Pups only acquired a subset of the adult syllable type repertoire (C) and adult syllable type acquisition followed a non-linear pattern (D). Bouts were dominated by reduplication of syllable types (E) and included syllable trains with regular beats (F). Babbling did not require a social context for production and the adult-like syllable types did not elicit the same reactions that they would when produced by adult bats (G). All pups – irrespective of sex and regional origin – engaged in babbling behavior (and showed no significant differences concerning the babbling features).

In conclusion, our study demonstrated that the babbling behavior of *S. bilineata* pups is characterized by the same features that define infant babbling (Fernandez et al., 2021).

The similarities in babbling features between two species with common traits such as VPL, laryngeal sound production and similar brain architecture, are a promising basis for comparative investigations of neuronal substrates in mammalian VPL (Jarvis, 2019).

## References

Fernandez, A. A., Burchardt, L. S., Nagy, M., & Knörnschild, M. (2021). Babbling in a vocal learning bat resembles human infant babbling. *Science, 373*(6557), 923-926. doi:10.1126/science.abf9279

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science, 366*(6461), 50-54. doi:10.1126/science.aax0287

Knörnschild, M., Behr, O., & von Helversen, O. (2006). Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). *Naturwissenschaften, 93*(9), 451-454. doi:10.1007/s00114-006-0127-9

Knörnschild, M., Nagy, M., Metz, M., Mayer, F., & von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters, 6*(2), 156-159. doi:10.1098/rsbl.2009.0685

Oller, D. K. (2000). *The Emergence of the Speech Capacity*: Mahwah, NJ, Lawrence Erlbaum Associates, Inc.

ter Haar, S. M., Fernandez, A. A., Gratier, M., Knörnschild, M., Levelt, C., Moore, R. K., . . . Oller, D. K. (2021). Cross-species parallels in babbling: animals and algorithms. *Philosophical Transactions of the Royal Society B: Biological Sciences, 376*(1836), 20200239. doi:10.1098/rstb.2020.0239

Vihman, M. M. (2014). *Phonological development: The first two years* (2 ed.): John Wiley and Sons, Inc.

# FROM PROTO-CONVERSATION TO MODERN CONVERSATION: THE ROLE OF HUMAN SELF-DOMESTICATION

FRANCESCO FERRETTI*[1], INES ADORNETTI[1], LJILJANA PROGOVAC[2],
ANTONIO BENÍTEZ BURRACO[3]

*Corresponding Author: francesco.ferretti@uniroma3.it
[1]Department of Philosophy, Communication and Performing Arts, Roma Tre University, Rome, Italy
[2] College of Liberal Arts & Science, Wayne State University, Detroit, USA
[3] Department of Spanish, Linguistics, and Theory of Literature (Linguistics), University of Seville, Seville, Spain

We propose that the evolution of one of the hallmarks of human communication, i.e., conversational abilities, is strongly linked to the reduction of reactive aggression, one of the triggering factors in human self-domestication (HSD), the hypothesis that humans evolved through changes that parallel those found in domesticated animals (Hare, 2017; Shilton et al., 2020). Within this framework, we focus on proto-conversation (PC) and proto-syntax at the basis of the communicative strategies of early anatomically-modern humans (AMH), viewing them as an evolutionary platform for evolving fully fledged forms of conversation and syntax.

Most influential models of the evolution of conversation in humans sharply distinguish proto-conversation (PC) and full-fledged conversation. For instance, Tomasello (2003; 2019) advocates an abrupt transition from animal PC to human conversation, that largely depends on the evolutionary emergence of cognitive strategies capable of ensuring the "triangulation" between interlocutors and objects of the outside world. This hypothesis builds on his analysis of the ontogenesis of human conversational abilities, and particularly, his view that an abrupt change takes place in the child at the 9th month, when face-to-face (dyadic) communication between mother and child is replaced by symbolic (triadic) communication. We instead argue in favor of a stronger continuity of human conversation and animal PC, as well as a smoother transition from PC to sophisticated conversational skills in infants. Arguing for a continuity between dyadic and triadic forms of intersubjectivity and ultimately, of PC and conversation, is in line with the view that higher-order competences are prepared and supported by a series of evolutionary steps (e.g. Trevarthen 1974; 1979; Bråten and Trevarthen, 2007), but also with the finding of turn-taking and associated rules of conversation in non-human animal vocal interactions (e.g., Pika et al., 2018), and more generally, with claims of an evolutionary continuity between human and animal communication (Levinson, 2016).

Our specific proposal is that the transition to full-fledged syntax and to full-fledged conversation is paralleled, and in fact facilitated by an increasing control of reactive aggression, which is a hallmark of the human behavioral phenotype and the ultimate reason of HSD. Importantly for our argument here, PCs are strongly connected to the emotional level (Delafield-Butt and Trevarthen, 2015). Accordingly, we contend that early stages of the dyadic mother-child relationship do not only represent the platform for the ontogenetic transition to triadic communication, but also a basic strategy for reassurance of the child's tension and distress, and ultimately, for controlling and reducing reactive aggression. This developmental aspect has implications for our model of language evolution under the effects of HSD. We propose that during phylogeny, the reduction of reactive aggression under the effects of HSD prompted the emergence of face-to-face communication that is a necessary prerequisite for conversational communication in humans. In turn, conversation triggered more complex forms of grammar that further contributed to transform physical competition into verbal competition. In this view, sophisticated uses of language, complexification of grammar, and reduction of aggression are involved in a process of coevolution - a feedback loop in which HSD is both the cause and the result of the enhancement of grammar and conversational pragmatics (Benítez-Burraco, Progovac, 2020; 2021; Benítez-Burraco et al., 2021).Within a framework of this kind, PC becomes a precondition for fully developed (symbolic) interactions, thus stressing the dialogic-persuasive character of communication rather than its symbolic-descriptive nature.

## References

Benítez-Burraco, A., & Progovac, L. (2021). Language evolution: examining the link between cross-modality and aggression through the lens of disorders. *Philosophical Transactions of the Royal Society B, 376*(1824), 20200188.

Benítez-Burraco, A., & Progovac, L. (2020). A four-stage model for language evolution under the effects of human self-domestication. *Language & Communication, 73,* 1–17.

Benítez-Burraco, A., Ferretti, F., & Progovac, L. (2021). Human Self-Domestication and the Evolution of Pragmatics. *Cognitive Science, 45*(6), e12987.

Braten, S., & Trevarthen, C. (2007). From infant intersubjectivity and participant movements to simulation and conversation in cultural common sense. In: *On being moved: From mirror neurons to empathy, 68,* 21.

Delafield-Butt, J. T., & Trevarthen, C. (2015). The ontogenesis of narrative: from moving to meaning. *Frontiers in psychology, 6*, 1157.

Hare B. (2017). Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual Review of Psychology, 68,* 155–186.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences, 20*(1), 6-14.

Pika, S., Wilkinson, R., Kendrick, K. H., & Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proceedings of the Royal Society B, 285*(1880), 20180598.

Shilton, D., Breski, M., Dor, D., & Jablonka, E. (2020). Human social evolution: self-domestication or self-control?. *Frontiers in Psychology, 11*, 134.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tomasello, M. (2019). *Becoming human*. Cambridge, MA: Harvard University Press.

Trevarthen, C. (1974). Conversation with a two-month-old. *New Scientist, 2*, 230–235.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In: M. Bullowa (Ed.), *Before Speech* (pp. 227–270). Cambridge: Cambridge University Press.

# LANGUAGE CONTACT IS PART OF LANGUAGE EMERGENCE: NOUN PHRASE ORDERING IN NICARAGUAN SIGN LANGUAGE

MOLLY FLAHERTY*[1], SAVITHRY NAMBOODIRIPAD[2], AND SUSAN GOLDIN-MEADOW[3]

*Corresponding Author: moflaherty@davidson.edu
[1]Department of Psychology, Davidson College, Davidson, USA
[2]Department of Linguistics, University of Michigan, Ann Arbor, USA
[3]Department of Psychology, The University of Chicago, Chicago, USA

Research on language emergence typically conceptualizes the birth of grammatical structure in a vacuum, i.e., an environment free of other languages. New languages, however, are not born in petri dishes. Here we consider a case of potential language contact between a new sign language in Nicaragua and the spoken language with which it cohabitates: Spanish.

We investigated noun phrase (NP) ordering in three successive cohorts in Nicaraguan Sign Language (NSL): Cohort 1, who came together in the 1970s and first formed NSL; Cohort 2 and Cohort 3, who were exposed to NSL upon school entry 1986-1998. We showed participants a series of cards depicting sets of objects (e.g., dogs, cars). Set size varied 1-4, and objects were either large or small. Participants described each card, and we classified their descriptions by the ordering of NP elements. Data was collected in 2009 (N=6) and 2015 (N=24, including the original 6).

Both lab-based artificial language learning studies (e.g., Culbertson et al., 2012) and cross linguistic data (see Dryer, 2008) indicate that adults prefer harmonic noun phrase ordering patterns, where adjective and number are expressed either both prenominally or both postnominally. In agreement with these findings, in 2009, we found that Cohort 1 and Cohort 2 signers produced harmonic orders (either noun-adjective-number or noun-number-adjective), with no significant difference between cohorts. In 2015, the pattern was the same for Cohort 1, who still preferred harmonic orders. Unexpectedly however, Cohort 2 signers increased their use of the non-harmonic number-noun-adjective order, which was also the preferred order for Cohort 3 signers. The preference for non-harmonic order increased significantly with cohort ($\beta$=-5.24, p<.02). This means in 2015, *individual signers* in the second cohort moved away from the harmonic pattern they displayed in 2009.

We propose language contact as an explanation for this unanticipated shift. This situation meets Thomason (2001)'s criteria for establishing contact-induced change: there is no language-internal motivation for this shift, word order is a highly borrowable linguistic feature (Bickel & Hickey, 2017; Thomason & Kaufman, 1988), and the order of nouns, numerals, and adjectives in NSL in 2015 matches Spanish order. The shift's rapidity has precedent in other dynamic contact contexts; Fortescue (1993) found Native Alaskan languages under pressure from English went from being flexible SOV to rigidly SVO within two generations.

We are currently in the preliminary stages of further investigating potential language contact between NSL and Spanish during the period 2009-2015. By interviewing community members about their experiences with written Spanish (in the form of text messaging and social media use) over this period, we will probe the degree to which Spanish use may have changed at the time of the shift. Should we find that signers' use of written Spanish increased during this period, this will strengthen our case for language contact. In addition, we hypothesize that differences in patterns of Spanish use could explain why Cohort 1 did not shift toward the Spanish order as Cohort 2 did.

By combining interviews about language use and ideology with experimental data, we draw on the approach advocated by Hou & de Vos (2021), who emphasize a need for situating longitudinal data on language structure with details about sociolinguistic contexts of learning and use. In this view, centering variation and dynamics is a way to move away from essentialist claims about particular languages. Further, contact between spoken and signed languages is an ideologically charged area. As communities fight for recognition and access to signed languages, there is substantial pressure to establish that signed languages are not just "Spanish on the hands" (e.g., Rhodes 2020). We argue this is due to deficit models towards multilingualism and language contact, which are present in society at large and also reflected in linguistics and language emergence research. Ignoring the multilingualism of signers, as well as the range of semiotic resources available to signers as they "language" across contexts (e.g., Moriarty Harrelson, 2019) does not serve to give an accurate picture of how conventions emerge and change in languages such as NSL (cf. Ansaldo, 2017). As language contact is the norm around the world, we seek to open a discussion of what we can gain by bringing language contact to the center of our models of language evolution.

## References

Ansaldo, U. (2017). Creole complexity in sociolinguistic perspective. *Language sciences*, *60*, 26-35.

Bickel, B., & Hickey, R. (2017). Areas and universals. *Cambridge Handbooks in Language and Linguistics*, 40-55.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306-329.

Dryer, M. S. (2008). Order of numeral and noun. *The world atlas of language structures online*, 362-365.

Fortescue, M. (1993). Eskimo word order variation and its contact-induced perturbation. *Journal of Linguistics*, *29*(2), 267-289.

Hou, L. & de Vos, C. (2021). Classifications and typologies: Labeling sign languages and signing communities. *Journal of Sociolinguistics*. 00:1-8. https://doi.org/10.1111/josl.12490

Moriarty Harrelson, E. (2019). Deaf people with "no language": Mobility and flexible accumulation in languaging practices of deaf people in Cambodia . *Applied Linguistics Review*, *10*(1), 55-72.

Rhodes, C. R. (2020). Dually authenticated and doubly modern: Institutionalizing jach maaya in the Yucatan today. *Journal of Linguistic Anthropology*, *30*(3), 326-352.

Thomason, S. G. (2001). *Language contact*. Edinburgh University Press.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Univ of California Press.

# SOCIAL TOLERANCE AND INTERACTIONAL OPPORTUNITIES AS DRIVERS OF GESTURAL REDOINGS IN ORANG-UTANS

MARLEN FRÖHLICH [*1,2], and CAREL VAN SCHAIK [2]

[*]Corresponding Author: marlen.froehlich@uni-tuebingen.de
[1]Paleoanthropology lab, Institute for Archaeological Sciences, University of Tübingen, Tübingen, Germany
[2]Department of Anthropology, University of Zurich, Zurich, Switzerland

Communicative repair is a fundamental and universal element of interactive language use. It has been suggested that the persistence and elaboration after communicative breakdown, as frequently observed in non-human primates, constitute evolutionary building blocks of this capacity [1], but the conditions favouring it are poorly understood. Social tolerance and interaction opportunities beyond the mother-offspring dyad may be central drivers of gestural redoings (i.e. repetition and/or modification of an initial signal after communicative failure) as a crucial element of social action coordination. Because zoo-housed individuals of some species are more sociable and terrestrial than in the wild, they should be more likely to produce and respond to gestural redoings during the coordination of joint activities.

We here examine this question by comparing the same species living in the wild and in artificial, man-made habitats in captivity. The wild-captive contrast allows us to directly test the prediction that captive individuals more readily exhibit persistence and elaboration during face-to-face interactions, because captivity's more social and terrestrial settings may foster the extensive use of redoings in the coordination of joint activities. We tested this prediction in orang-utans, a great ape genus which is in our view ideal for this avenue of research. First, the orang-utan populations of Borneo (*Pongo pygmaeus wurmbii*) and Northwest-Sumatra (i.e. Suaq and Ketambe, *Pongo abelii*) differ considerably in sociability [2]) and social tolerance (Bornean orang-utans become more stressed in group settings than Sumatrans [3]). Second, in contrast to natural environments, captive orang-utans are always in close proximity and more on the ground, clearing sight and freeing hands for visual and tactile communication. Third, in addition to these setting and taxonomic contrasts, the pairing of social partners (i.e. the interaction dyad) also affects features of social interactions, e.g. due to differences in social tolerance and familiarity. Our recent work demonstrated remarkable behavioural plasticity in orang-utans, with regard to communicative repertoires [4] and multimodal use of communicative acts contingent on social context and partner [5]. There are no systematic wild-captive comparisons of apes' communicative behaviour to date, but we assume that contrasts must be larger for orang-utans than any other great ape taxon in light of the profound contrasts in sociality and arboreality between captivity and the wild.

We studied a large comparative sample (N = 3869 signal instances) of wild and zoo-housed orang-utans of two different species (*Pongo abelii, P. pygmaeus*), focusing on the gestural solicitation of three distinct joint activities: social play, allo-grooming and joint travel. Specifically, we tested how research setting, species and interaction dyad affected the probability of repetition and elaboration in signal use after communicative failure, as well as the success in obtaining the original goal (the "apparently satisfactory outcome"; [6]), while controlling for critical individual variables and context. Our predictions were confirmed only for elaboration, the more flexible form of redoings. Specifically, results showed that gestural redoings in general were best predicted by the specific social context (i.e. social play) and interaction dyad (i.e. beyond mother-offspring), although were least frequent in captive Bornean orang-utans. For gestural elaboration, we found the expected differences between captive and wild research settings in Borneans, but not in Sumatrans (the more socially tolerant species). Moreover, we found that the effectiveness of elaboration in eliciting responses was higher in Sumatrans, especially the captive ones, whereas effectiveness of mere repetition was influenced by neither species nor setting.

In sum, this study demonstrated that Bornean and Sumatran orang-utans living in different socio-ecological environments (i.e. wild versus captive settings) frequently deploy gestural redoings in the face of communicative failure within and beyond mother-offspring dyads. Both repetition and elaboration have been interpreted as simple forms of repair [1], a fundamental layer of the human interaction engine [7]. Our findings also showed that social tolerance, as a foundation for extended social interactions (both higher in Sumatrans and in captivity), plays a central role in the emergence of complex exchanges in great apes. These findings support the notion that the human version of the interaction engine evolved when contexts with limited conflicts of interests and involving coordination for joint action became more widespread [8]. The Interdependence Hypothesis further posits that as individuals in hominin societies became more cohesive and interdependent, pressure for shared intentionality, and thus effective coordination via communication increased [8]. As parts of the interactional base for joint action coordination, this includes the repair of communicative "glitches" in case of misunderstandings as well as negotiation in case of diverging interests [7]. For future empirical research on nonhuman primates, it would be important to distinguish communicative sequences resulting from prior communicative failure alone (and thus primarily driven by the signaller) from more interactional communicative exchanges involving signalling by both parties, as common in human conversation.

## References

1.Heesen R., Fröhlich M., Sievers C., Woensdregt M., Dingemanse M. (in press) Coordinating social action: A primer for the cross-species investigation of communicative repair *Philosophical Transactions of the Royal Society B: Biological Sciences*.

2.van Schaik C.P. (1999) The socioecology of fission-fusion sociality in orangutans. *Primates* **40**(1), 69-86.

3.Weingrill T., Willems E.P., Zimmermann N., Steinmetz H., Heistermann M. (2011) Species-specific patterns in fecal glucocorticoid and androgen levels in zoo-living orangutans (*Pongo* spp.). *General and Comparative Endocrinology* **172**(3), 446-457.

4.Fröhlich M., Bartolotta N., Fryns C., Wagner C., Momon L., Jaffrezic M., Mitra Setia T., Schuppli C., Noordwijk M.A., van Schaik Carel P. (2021) Orangutans have larger gestural repertoires in captivity than in the wild – a case of weak innovation? *iScience* **24**(11), 103304.

5.Fröhlich M., Bartolotta N., Fryns C., Wagner C., Momon L., Jaffrezic M., Noordwijk M.A., van Schaik C.P. (2021) Multicomponent and multisensory communicative acts in orang-utans may serve different functions. *Communications Biology* **4**, 917.

6.Hobaiter C., Byrne R. (2014) The meanings of chimpanzee gestures. *Current Biology* **24**(14), 1596-1600.

7.Levinson S.C. 2019 Interactional foundations of language: the interaction engine hypothesis. In *Human language: From genes and brain to behavior* (ed. Hagoort P.), pp. 189-200. Cambridge, US, MIT Press.

8.Tomasello M., Melis A.P., Tennie C., Wyman E., Herrmann E. (2012) Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis. *Current Anthropology* **53**(6), 673-692.

# PRIMATE RHYTHMIC CATEGORIES ANALYZED ON AN INDIVIDUAL BASIS

MARCO GAMBA[1], CHIARA DE GREGORIO[1], DARIA VALENTE[1], TERESA RAIMONDI[1], VALERIA TORTI[1], LONGONDRAZA MIARETSOA[1], FILIPPO CARUGATI[1], OLIVIER FRIARD[1], CRISTINA GIACOMA[1], ANDREA RAVIGNANI[2,3]

*Corresponding Author: marco.gamba@unito.it
[1]Department of Life Sciences and Systems Biology, University of Torino, Torino, Italy
[2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[3]Aarhus University, Aarhus, Denmark

Rhythm is a fundamental feature characterizing communicative displays, and recent studies showed that primate songs encompass categorical rhythms falling on small integer ratios observed in humans. We individually assessed the presence and sexual dimorphism of rhythmic categories, analyzing songs emitted by 39 wild indris. Considering the intervals between the units given during each song, we extracted 13556 interval ratios and found three peaks (at around 0.33, 0.47, and 0.70). Two peaks indicated rhythmic categories corresponding to small integer ratios (1:1, 2:1). All individuals showed a peak at 0.70, and most showed those at 0.47 and 0.33. In addition, we found sex differences in the peak at 0.47 only, with males showing lower values than females. This work investigates the presence of individual rhythmic categories in a non-human species; further research may highlight the significance of rhythmicity and untie selective pressures that guided its evolution across species, including humans.

## 1. Introduction

Rhythm is a fundamental feature of human communicative displays and characteristically permeates the musical performances of our species (Hausen et al. 2013, Savage et al. 2015). When we think of a typical human communicative display, such as conversation, we realize that numerous factors play a critical role in the temporality of exchanges between interlocutors (Pouw & Holler 2022). Among these features is rhythm, which is often contextually adapted to align semantic and syntactic processes (Himberg et al. 2015).

Moreover, we know that complex rhythms characterize communication in human and non-human animals (Pouw et al. 2021; Ravignani et al. 2019). Recent research also indicated comparable features in the temporal structure characterizing human speech, music, and non-human animal vocalizations (Kello et al. 2017; Pouw et al. 2021). Complex rhythms are typical of interactions between two or more conspecifics, emitting a series of signals and taking turns (Gamba et al. 2011; 2016). It is difficult to imagine that we will be able to discern the cognitive processes involved in turn-taking for a large number of species, but finding similarities across extant primates could help us to understand the selective pressures that may have played a role during the evolution of human language (Levinson 2016).

Indeed, rhythmic displays can provide important information about the communicative features of other species, in particular concerning the contextual variation and temporal characteristics that permeate long sequences of emissions (Clink et al. 2017). Specifically, recent research has shown that the study of primate songs may indicate the presence of rhythmic categories that fall on small integer ratios which at least partially overlap with those we can observe in human music (De Gregorio et al. 2021).

A question remains unanswered: to what extent rhythmic patterns are stable across the individuals of a particular species and whether all individuals share a certain rhythmic category more often than others. In order to shed light on this aspect, we investigated the acoustic emissions of a singing primate that has already been the subject of studies on rhythmic categories, *Indri indri*. Indri is a primate species with an evolutionary history parted from humans by around 74 million years (Hedges et al. 2015). Our first hypothesis is that adult individuals of this species present at least two rhythmic categories in their songs. We, therefore, expect that the density distributions of all studied indris show two or three peaks around human typical small integer ratios of 1:2 (0.33), 1:1 (0.50), and 2:1 (0.67), following the results of previous work (De Gregorio et al. 2021). The second hypothesis, alternative to the previous one, is that rhythmic categories appear at least consistently within one of the sexes, in line with the fact that, in the indri, many of the vocal features are distinctly dimorphic between males and females (Giacoma et al. 2010; Gamba et al. 2016; De Gregorio et al. 2019; Zanoli et al. 2020, Valente et al. 2021).

## 2. Materials and Methods

### *2.1. Study subjects and recordings*

We recorded songs produced by 39 reproductive indris of 20 different groups. As indri groups usually comprise a reproductive pair and their offspring, by doing so we included in the analyses a number of individuals comparable for each of the sampled group. These indris lived in various forest sites in Madagascar: Analamazaotra Reserve (Andasibe-Mantadia

National Park), Mantadia (Andasibe-Mantadia National Park), Mitsinjo Station Forestière, Maromizaha Forest.

We collected data between 2005 and 2016 using solid-state recorders (SoundDevices 702, Olympus S100 and LS05, and Tascam DR-100, DR-40, and DR-05) equipped with Sennheiser (ME 66 and ME 67) or AKG (CK 98) microphones. We set a sampling rate of 44.1 kHz and 16-bit amplitude resolution for all recording sessions. The distance ranged from 2 to 20 meters when recording the animals, with the microphone aimed at the focal singing animal.

### *2.2. Acoustic and statistical analyses*

We tracked inter onset intervals between the units given during each song using Praat TextGrids, where we assigned each singer to a particular TextGrid tier. After extracting interval duration to a text file (Gamba et al. 2016), we obtained 13556 interval ratios using the Roeske et al. (2020) formula. Interval ratios $r_k$ were calculated by dividing a particular inter-onset interval $t_k$ for itself + the following inter-onset interval ($r_k = t_k /(t_k+t_{k+1})$).

## 3. Results

The distribution of the peaks in the $r_k$ density graphs showed three different patterns (Fig. 1A). A first pattern, occurring in 56% of the individuals, showed the presence of three peaks means at 0.33 ± 0.02, 0.47 ± 0.02, 0.70 ± 0.02. All individuals had a peak at around 0.70 (Fig. 1B). In 26% of the indris, the peak at around 0.33 was missing, while 18% did not show the peak at around 0.47 (Fig. 1B).

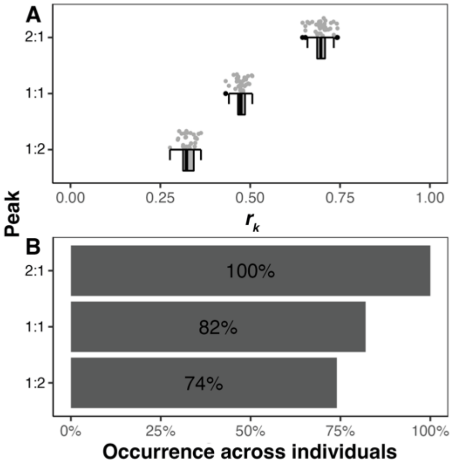


Figure 1. Variation of the rhythm ratios ($r_k$) across individual indris for each peak of their density plot (A, see also Fig. 2). Grey points denote individual rhythm ratios for a particular peak, black points denote outliers. Percentage distribution of rhythm ratios across individuals for the three peaks (B, see also Fig. 2).

The coefficient of variation of the peak around 0.33 was 6.1%, while the other two showed lower values, 3.5% and 3.4%, respectively (Fig. 1A, Fig. 2). We analysed the occurrence of peaks across the sexes and found no statistical difference for the peak at 0.70 (Mann-Whitney test, W = 213, p-value = 0.5315). The peak at 0.33 showed marginal significance, but small effect size (W = 57, p-value = 0.0476, r = -0.44), whereas the peak at around 0.47 showed significant difference (W = 188, p-value = 0.02195, r = 0.47). Males showed an average peak of 0.47±0.02, while females showed values at 0.48±0.01.

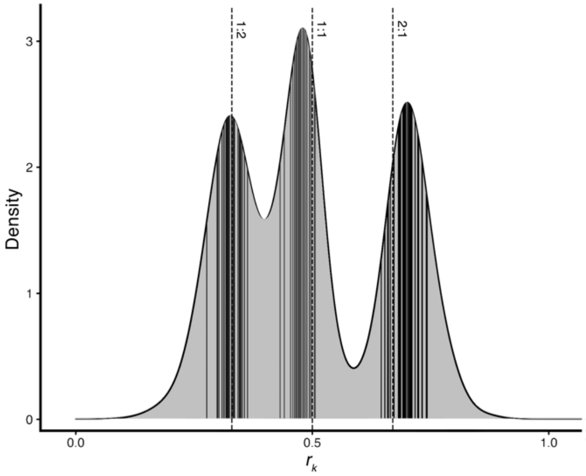


Figure 2. Probability density function of rhythm ratios ($r_k$), which we calculated across 39 adult indris and 636 individual contributions to songs. Solid lines denote peaks of each individual indri for each group of peaks. Dotted lines denote 1:2 (0.33), 1:1 (0.50), and 2:1 (0.67) ratios.

## 4. Discussion

In agreement with the De Gregorio and colleagues (2021) study, we observed that most indri part of our sample showed three distinct peaks representing as many rhythmic categories. The location of the peaks is in agreement with the two rhythmic categories, significantly corresponding to small integer ratios in De Gregorio and colleagues (2021). The peak at 2:1 (0.67) was not statistically significant in the previously cited work because points around 0.67 were not dense enough. Although the methodology used here is slightly different, we see the same trend, showing a well-identified and consistent peak at 0.70+0.02 in most animals studied. In most individuals, we observe three peaks; however, the 44% of indris shows only two peaks, distributed over two of the three categories mentioned above. Indeed, the presence of two and not three rhythmic categories is not easily attributable to the sex of the individuals. Interestingly, the absence of the third peak does not alter the consistency of the remainder concerning the species-specific characteristics that emerged in this study and De Gregorio and colleagues (2021).

De Gregorio and colleagues' (2021) findings concerning male and female absolute interval lengths are different, but the 1:1 ratios are the same and appear consistent with our overall results. However, by analyzing individual peaks, a slight difference emerges for values below 1:1 (i.e., isochrony).

Indri groups are often small, include a pair and their offspring, and communicate in a very extensive network of other conspecific groups. Unlike what has been suggested for birdsong, which undergoes a vocal learning process, we do not know what role learning plays in primate songs. Historically, it has been assumed that singing has a strong genetic component, but recent studies point to plasticity that is difficult to explain by factors that do not involve some degree of learning.

This work is the first to analyze rhythmic categories at the individual level because previous investigations have focused on a species-specific pattern rather than an individual pattern (Roeske et al. 2020; De Gregorio et al., 2021). Given the hierarchical organization of natural systems, expanding the levels of analyses can provide insight regarding the evolutionary levels on which selection mechanisms may act. Our results suggest new insights for subsequent research, which may be aimed at understanding whether the presence of rhythmic categories varies with the age of individuals, or is a trait that becomes fixed early and then does not change during ontogeny, or whether it varies dynamically over time. Our study also shows that most indris studied exhibit rhythmic categories that correspond to small integer ratios. Unfortunately, our findings concerning the production of categorical rhythmic patterns in the indris cannot be paralleled by pieces of evidence about the perception of these lemurs. However, it appears challenging to study whether the indris' perceptual abilities are biased towards the same categories in the wild.

## Acknowledgements

This research was supported by Università degli Studi di Torino and the African, Caribbean, and Pacific (ACP) Science and Technology Programme of the ACP Group of States, with the financial assistance of the European Union and by grants from the Parco Natura Viva—Centro Tutela Specie Minacciate. We are grateful to GERP (Groupe d'Etude et des Recherche sur les Primates) and Dr. Jonah Ratsimbazafy. We thank Dr. Cesare Avesani Zaborra and Dr. Caterina Spiezio for helping us with the organization of the field station in Maromizaha. We are grateful to the researchers and the research guides, to Mamatin and all our Malagasy collaborators in the forest, for their help and logistical support. The contents of this document are the sole responsibility of the authors and can under no circumstances be regarded as reflecting the position of the European Union. We have received permits for all the research activities from “Direction des Eaux et Forêts” and “Madagascar National Parks” (formerly ANGAP).

## References

Clink, D. J., Bernard, H., Crofoot, M. C., & Marshall, A. J. (2017). Investigating individual vocal signatures and small-scale patterns of geographic variation in female Bornean gibbon (*Hylobates muelleri*) great calls. *International Journal of Primatology*, 38(4), 656-671.

De Gregorio, C., Zanoli, A., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., Giacoma, C. & Gamba, M. (2019). Female indris determine the rhythmic structure of the song and sustain a higher cost when the chorus size increases. *Current Zoology*, 65(1), 89-97.

Gamba, M., Favaro, L., Torti, V., Sorrentino, V., & Giacoma, C. (2011). Vocal tract flexibility and variation in the vocal output in wild indris. *Bioacoustics*, 20(3), 251-265.

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., Bonadonna, G., Friard, O. & Giacoma, C. (2016). The indris have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. *Frontiers in Neuroscience*, 10, 249.

Giacoma, C., Sorrentino, V., Rabarivola, C., & Gamba, M. (2010). Sex differences in the song of *Indri indri*. *International Journal of Primatology*, 31(4), 539-551.

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: a common rhythm. *Frontiers in psychology*, 4, 566.

Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. *Molecular biology and evolution*, 32(4), 835-845.

Himberg, T., Hirvenkari, L., Mandel, A., & Hari, R. (2015). Word-by-word entrainment of speech rhythm during joint story building. *Frontiers in Psychology*, 6, 797.

Kello C.T., Bella S.D., Médé B., Balasubramaniam R. (2017) Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of the Royal Society Interface* 14

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, 20(1), 6-14.

Pouw, W., & Holler, J. (2022). Timing in conversation is dynamically adjusted turn by turn in dyadic telephone conversations. *Cognition*, 222, 105015.

Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R.S. & Wiggins, G. A. (2021). Multilevel rhythms in multimodal communication. *Philosophical Transactions of the Royal Society B*, 376(1835), 20200334.

Ravignani, A., Verga, L., & Greenfield, M. D. (2019). Interactive rhythms across species: the evolutionary biology of animal chorusing and turn-taking. *Annals of the New York Academy of Sciences*, 1453(1), 12-21.

Roeske, T. C., Tchernichovski, O., Poeppel, D., & Jacoby, N. (2020). Categorical rhythms are shared between songbirds and humans. *Current Biology*, 30(18), 3544-3555.

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. Proceedings of the National Academy of Science, 112(29), 8987-8992.

Valente, D., De Gregorio, C., Favaro, L., Friard, O., Miaretsoa, L., Raimondi, T., Ratsimbazafy, J., Torti, V., Zanoli, A., Giacoma, C., & Gamba, M. (2021). Linguistic laws of brevity: conformity in *Indri indri*. *Animal cognition*, 24(4), 897-906.

Zanoli, A., De Gregorio, C., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., Giacoma, C., & Gamba, M. (2020). Sexually dimorphic phrase organization in the song of the indris (*Indri indri*). *American Journal of Primatology*, 82(6), e23132.

# CONSTRAINING FREE MERGE: LABELING AND THE THETA-CRITERION

JASON GINSBURG*1

*Corresponding Author: jginsbur@cc.osaka-kyoiku.ac.jp
1Department of Educational Collaboration, Osaka Kyoiku University, Japan

Merge, the ability to recursively combine two elements together, is a core property of human language, and thus it is reasonable to assume that language evolved to make use of Merge. If Merge is essentially a free operation that the language module makes use of, it must be constrained in some ways. I created a computer model of language that implements Free Merge, and I explain how two properties of language, the need for syntactic structures to be labeled (Labeling), and constraints regarding theta assignment (the theta-criterion) can constrain Free Merge. This work suggests that labeling and the theta-criterion play an important role in language generation.

## 1. Introduction

Chomsky (2005, p. 11) writes that “the human language faculty is … a system of discrete infinity” that is “based on a primitive operation that takes *n* objects already constructed, and constructs from them a new object.” This process of combining together linguistic objects is referred to as Merge. Chomsky (2011, p. 52) writes that “unbounded Merge is the sole recursive operation within UG” and that it is “part of the genetic component of the language faculty.” If this is correct, human language evolved to make use of recursive Merge.[1]

Chomsky (2001, 2013, 2015) takes the position that Merge is free. Chomsky (2015, p. 14) writes that “[t]he simplest conclusion … would be that Merge applies freely” and “[o]perations can be free, with the outcome evaluated at the phase level for transfer and interpretation at the interfaces.” Free Merge would result in an infinite number of possible structures generated for every possible utterance. Thus, Free Merge must be constrained by the language faculty.

---

[1] Berwick (2011) suggests that non-human primates have lexical items but no Merge, whereas birds have something like Merge (used in songs) but no lexical items. Human language, crucially, makes use of lexical items and Merge.

I utilized a computer model that automatically generates complete derivations of sentences to test the limits of Free Merge. With this model, I was able to constrain Free Merge with two language-related principles utilized in Minimalist work, Labeling and the theta-criterion.

## 2. Merge

Merge involves combining together syntactic objects (SOs). External Merge refers to selecting a lexical item from a workspace and Merging it with another lexical item or SO, which is an already formed syntactic structure. Internal Merge refers to selecting an element from within an SO and Merging it again with the root of the SO.

There are two types of Merge: set-Merge and pair-Merge. Figure 1a shows set-Merge of the two SOs X and Y. In this case, X is the label. Figure 1b shows Z pair-Merged with an XP; pair-Merge is represented as a dotted arc. Crucially, Z could be considered an adjunct on a separate plane.

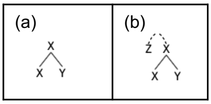


Figure 1. Set-Merge and Pair-Merge

Given Free Merge, it is possible to internally set-Merge an argument freely. In the derivation of *Tom read a book* (Chomsky, 2015), assume that LIs are selected from an input list in (1)a, and that at each stage of a derivation, an LI can be selected from the Input List and externally set-Merged with an already formed SO, or an SO can be internally set-Merged. The desired output is (1)b, following Chomsky's (2015) view that an object internally set-Merges with the verbal root and the verbal root *read* raises to v*. The subject also raises from the v*P to the TP. Note that Free Merge would also produce the ill-formed structures in (1)c-d, in which *a book* is internally set-Merged multiple times, and ends up in positions in which it is not found in actual language, so these types of structures must crash.

(1) (a) Input list: [C, T, Tom, v*, read, a, book]
(b) C Tom T ~~Tom~~ v*+read a book ~~read a book~~. → Tom read a book.
(c) C a book T Tom v*+read a book ~~read a book~~. → *A book Tom read.
(d) C Tom T a book ~~Tom~~ v*+read ~~a book read a book~~. → *Tom a book read.

The verb *think*, as in (2), differs from a transitive verb such as *read* in that it lacks case and doesn't agree with an object. To deal with this, I assume that the intransitive v (cf. Chomsky 2001) Merges with *think* and that v is able to label by itself.[2]

(2) He v thinks that John read the book.

## 3. Computer Model Output and Free Merge

I tested Free Merge using a computer model (which I constructed primarily in the Python programming language) that selects and Merges together lexical items from an input list, and automatically constructs detailed step-by-step derivations of phrases and sentences. This model incorporates Free Merge so that an argument can be freely internally set-Merged. This model also incorporates other core notions of Labeling Theory, such as feature transmission; a phase head (C or v*) passes unchecked phi-features to a complement and these unchecked phi-features need to form an Agree relation and be checked.

### 3.2 Labeling

In Labeling Theory (Chomsky, 2013, 2015), a label refers to the ability to interpret a phrase, e.g., as a VP, NP, etc. If a phrase cannot be labeled, then it cannot be properly interpreted. Murphy (2015) writes that "Label constitutes the evolutionary novelty which distinguishes the human cognome from non-human cognomes (7)", where the human cognome is "the finite set of operations available to the human brain (2)." Whether or not labeling is the core element that distinguishes human language from other animal communication systems remains an open question, but labeling appears to be crucial for human language.

The basic assumptions of Labeling Theory are that in an {H, XP} configuration in which H is a head and XP is a phrase, H can label (if it is strong enough to label). In an {XP, YP} configuration, if the heads of the XP and YP share prominent features, the shared feature(s) can label, or if one of the phrases moves out (undergoes internal Merge), the head of the remaining phrase can label

---

[2] Epstein, Kitahara, and Seely (2016) propose that *think* and v* are externally pair-Merged together, resulting in <think, v*>, where v* is the pair-Merged element. External pair-Merge of v* with *think* results in the uPhi, as well as the phasehood, of v* disappearing. This external pair-Merge analysis is one way to account for verbs such as *think*. However, permitting external pair-Merge greatly increases the possibilities for Free Merge. It also isn't entirely clear how phasehood and uPhi of a phase head can disappear under pair-Merge.

(if that head is strong enough to label).[3] In addition, Chomsky proposes that the English T and verbal roots are inherently too weak to label by themselves. However, when they Agree with an XP in an {XP, YP} structure, shared features strengthen them so that they can label.[4]

Consider the derivation of *Tom will read a book*, shown in Figure 2, which is successfully generated in my model. The object *a book*[5] internally set-Merges with the SO headed by *read*. This results in an {XP, YP} structure that is labeled via shared phi-features, indicated as <ϕ, ϕ>, resulting from Agree between *read* and *a book*. Similarly, the subject *Tom* internally set-Merges with the TP, resulting in a labelable {XP, YP} structure, resulting from the shared phi-features of the subject and T. Note that I follow Chomsky's (2015) view that the verbal root *read* moves to v*, producing <v*,read>, with v* the pair-Merged element. In addition, *will* moves to T forming <Tpres, will>. This "head-movement" is implemented as internal pair-Merge.

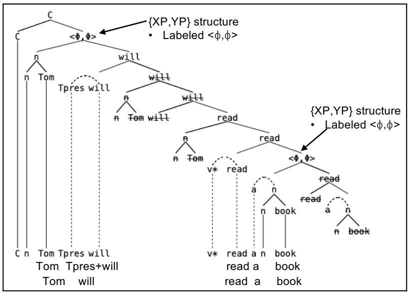


Figure 2. Successful derivation of *Tom will read a book*

---

[3] An anonymous reviewer asks about cases in which both elements in an {XP, YP} structure move out, such as in a French v*P where the subject and verb supposedly move. Movement of the verb is head movement. How exactly to deal with head movement is an open question. Chomsky (2001) suggests that head movement is a phonological operation. If head movement occurs at PF (phonological form), then head movement of the verb does not influence the syntax and will not influence labeling. In my model, the head is visible for labeling in its base position even after it has undergone movement, which I have implemented as internal pair-Merge.

[4] Murphy & Shim (2020) argue that T can label by itself. If uPhi are responsible for blocking labeling (cf. Mizuguchi 2017 for a proposal that heads without unvalued features can label), then it is possible that finite T (which has uPhi) cannot label by itself, whereas non-finite T (which lacks uPhi) is able to label. Murphy & Shim also note potential problems for permitting shared features to label, and they propose that only categorial heads label. Although these issues are worthy of further analysis, I assume the standard Labeling view that shared prominent features can label.

[5] Note that I assume that *a book* is an NP, with the determiner being a pair-Merged adjunct, a view that has been suggested by Chomsky (2007), as well as Oishi (2015). See Bruening (2020), among others, for support for the view that arguments are NPs and not DPs. Whether or not an argument is treated as an NP or a DP is not crucial for the examples discussed in this paper.

My model implements Free Merge of arguments, which can result in derivations that crash. In Figure 3, the subject *Tom* remains within the v*P and does not move to the specifier of TP. Since Merge is free, nothing forces the subject to move. However, this derivation crashes because of labeling failures, shown in Figure 3b. In the v*P, there is an {XP,YP} structure in which the subject and the v*P do not share phi-features, so labeling is not possible. In addition, assuming that English T is too weak to label by itself (following Chomsky 2015), T, represented as <Tpres, will>, is unable to label since it has no subject with which it shares phi-features. In another unsuccessful derivation shown in Figure 4, the object, *a book*, undergoes internal set-Merge with the TP, and the underlying subject *Tom* remains in-situ. This results in the labeling failures shown in Figure 4b. Although *a book* and <Tpres, will> are able to form a labelable {XP, YP} structure due to phi-feature agreement, the v*P-internal subject and the v*P form an unlabelable {XP, YP} structure, due to a lack of shared phi-features. Furthermore, the root *read* is also unlableable, since the object *the book* has moved out and the root *read* cannot be strengthened by shared phi-features.

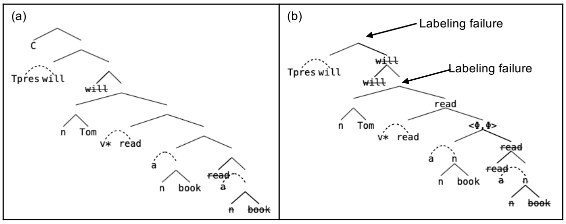


Figure 3. Labeling failure: unsuccessful derivation of *Tom will read a book*

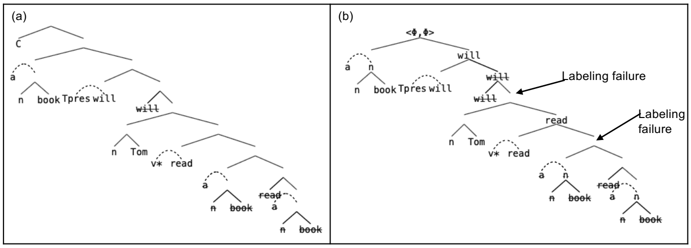


Figure 4. Labeling failure: another unsuccessful derivation of *Tom will read a book*

The sentence *He thinks that John read the book* has the successful derivation shown in Figure 5, which contains v that Merges with *think*, and v is able to label.

Free Merge also permits derivations such as that in Figure 6, in which the subject remains in-situ. This will crash because the matrix subject is in an unlabeleable {XP, YP} configuration with the vP and because T is not strengthened. Thus, this ill-formed derivation is ruled out by Labeling.

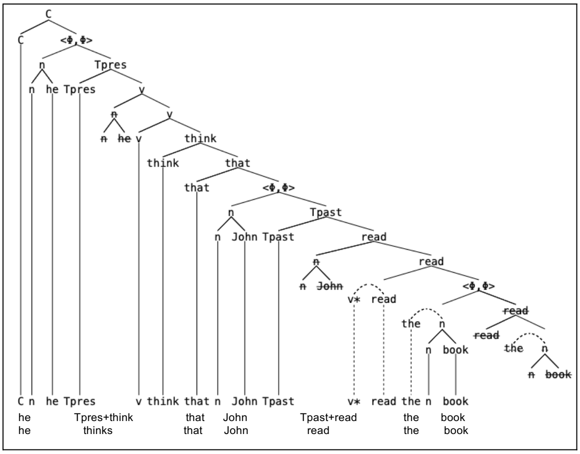


Figure 5. Successful derivation of *He thinks that John read the book*

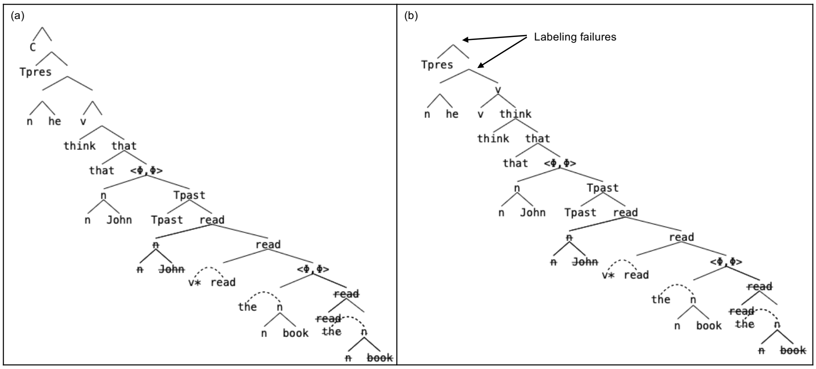


Figure 6. Labeling failure in *He thinks that John read the book*

### 3.3 Theta-Criterion

The theta-criterion is the notion that "[e]ach argument bears one and only one θ-role, and each θ-role is assigned to one and only one argument (Chomsky, 1981, p. 36)." The theta-criterion plays an important part in constraining Free Merge.

Figure 7 below shows an unsuccessful portion of a derivation, automatically generated by my model, of *Tom will read a book* (compare with the successful

Figure 2 above). The object *a book* is externally set-Merged with *read*, where it obtains a theta-role. The object then internally set-Merges with *read* and again with <v*, read> - this is the traditional specifier of v*P subject position. Since Merge is free, this operation is permitted. However, the transitive <v*, read> has a subject theta-role to assign. Thus, *a book* will get a second theta-role. This violates the theta-criterion, so this derivation crashes.

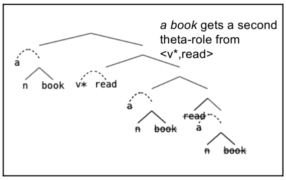


Figure 7. Theta-criterion violation: Unsuccessful derivation of *Tom will read a book*

## 4. Conclusion

I have discussed the derivations of 2 simple sentences and attempted to show how Labeling and the theta-criterion are sufficient for constraining Free Merge. The number of crashed and successful derivations[6] for these 2 sentences are summarized in Table 1. All of the crashes are caused by labeling failures or violations of the theta-criterion. Thus, labeling and the theta-criterion (in some form or other) seem to be core properties of language, and accounts of language evolution likely need to explain these properties.

Table 1. Crashes and successful derivations

| | Sentence | # of Crashes | # of Successful Derivations |
|---|---|---|---|
| (a) | *Tom will read a book.* | 14 | 2 |
| (b) | *He thinks that John read the book.* | 22 | 1 |

## 5. Supplementary Materials

The complete derivations for the target examples discussed in this paper can be found at: http://www.osaka-kyoiku.ac.jp/~jginsbur/JCole22Appendix

---

[6] In some cases, my model predicts that there can be more than 1 convergent derivation for a single sentence, as with *Tom will read a book* (which differ with respect to the number of times that the subject *Tom* undergoes internal set-Merge). This raises the question of whether a single well-formed sentence can have multiple possible structures resulting in the same interpretation.

## Acknowledgements

This research was supported by Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research #20K00664. I would like to thank the anonymous reviewers and Hiroshi Terada for helpful comments. All errors are my own.

## References

Berwick, R. (2011). All you need is Merge: Biology, computationa, and language from the bottom up. In A. M. Di Sciullo and C. Boeckx (Eds.), *The Biolinguistic Enterprise: New perspectives on the evolution and nature of the human language faculty* (pp. 461-491). Oxford, Oxford University Press.

Bruening, B. (2020). The head of the nominal is N, not D: N-to-D Movement, Hybrid Agreement, and conventionalized expressions. *Glossa: a journal of general linguistics,* 5(1) 15, 1–19.

Chomsky, Noam. (1981). *Lectures on government and binding: the Pisa lectures*. Berlin: Mouton de Gruyter.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: a life in language* (pp. 1-52). Cambridge, MA: MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36, 1-22.

Chomsky, N. (2007). Approaching UG from below. In U. Sauerland and H.-M. Gärtner (eds.), *Interfaces + Recursion = Language?* Chomsky's Minimalism and the view from syntax-semantics (pp. 1-30). The Hague: Mouton de Gruyter.

Chomsky, Noam. (2010). Some simple evo devo theses: How true might they be for language? In R. K. Larson, V. Déprez, and H. Yamakido (Eds.), *The evolution of human language* (pp. 45-62). Cambridge, UK: Cambridge University Press.

Chomsky, N. (2013). Problems of projection. *Lingua*, 130, 33-49.

Chomsky, N. (2015). Problems of projections: Extensions. In E. Di Domenico, C. Hamann, and S. Matteini (Eds.), *Structures, strategies and beyond: Studies in honour of Adriana Belletti* (pp. 1-16). Amsterdam: John Benjamins.

Epstein, S. D., Hisatusugu K., & Seely, T. D. (2016). Phase cancellation by external pair-merge of heads. *The Linguistic Review*, 33, 87-102.

Mizuguchi, M. (2017). Labelability and interpretability. *Studies in Generative Grammar* 27. 327-365.

Murphy, E. (2015). Labels, cognomes, and cyclic computation: An ethological perspective. *Frontiers in Psychology*, 6, 715.

Murphy, E. & Shim, J.-Y. (2020). Copy invisibility and (non)-categorial labeling. *Linguistic Research*, 37, 187-215.

Oishi, Masayuki. 2015. The hunt for a label. In H. Egashira et al. (Eds.), *Untiring pursuit of better alternatives* (pp. 222-334). Tokyo: Kaitakusha.

# GREAT APE PRAGMATICS

KIRSTY E. GRAHAM*[1], and CATHERINE HOBAITER 2[1]

*Corresponding Author: keg4@st-andrews.ac.uk
[1]School of Psychology & Neuroscience, University of St Andrews, UK

Increasingly, animal behaviour researchers are applying a linguistic framework to studying nonhuman communication (Berthet et al., 2021; Fishbein et al., 2020). While animal communication systems alone likely do not provide the key to human language evolution (Fitch, 2020), applying linguistic tools to communication systems beyond language can be revealing. Ethologists have long been aware of the impact of context on behaviour (e.g. Smith, 1965), and so it is somewhat surprising that pragmatics (the role of context in communication) has not been more widely applied to research on primate communication. Part of the problem has been that historically researchers have methodologically muddled semantics and pragmatics, allowing context to stand in place of meaning without considering the relationship between them. To disentangle the two, it is necessary to look separately at (1) the behavioural outcome of the signal that seems to satisfy the signaller, and (2) the context in which the signal was deployed.

Recent research in ape gestural communication has made great strides in semantics, determining the meanings of gestures independently of the immediate behavioural context or social relationship between signaler and recipient (Cartmill & Byrne, 2010; Graham et al., 2018; Hobaiter & Byrne, 2014). Pragmatics has received more attention in primate vocal research (Seyfarth & Cheney, 2018), although here the underlying intended meanings are not always clear. Given its large, diverse repertoires, used by all apes to achieve their day-to-day goals across contexts and with flexible meanings, gestural communication is a promising system for studying pragmatics in nonhuman species. And indeed it seems that bonobo gestures have different meanings in different contexts (Graham et al., 2020). We expect context to similarly impact meanings for other ape gestural communication systems, raising intriguing questions about the underlying cognitive mechanisms involved, and their shared similarities with those underpinning human language.

Here, we make a theoretical case for expanding the study of pragmatics across animal communication research. We sift through existing research on nonhuman communication to extract pragmatics where possible, and use this as a jumping off point from which to offer suggestions on ways of studying pragmatics in nonhuman species. We make specific recommendations in regards to recording and extracting data on different aspects of context that may be relevant to nonhuman interlocutors, such as behavioural (e.g. prior behaviour of signaller, recipient, or audience), social (e.g. kin relationships, rank), and environmental (e.g. vegetation density, weather) contexts, including how we may operationalise these. We draw on linguistic approaches and incorporate these with our own expertise in great ape gestural communication. We propose that broadening our view to include the context in which signals are deployed will provide new insight into understanding nonhuman communication from their perspectives.

**References**

Berthet, M., Coye, C., Dezecache, G., & Kuhn, J. (2021). Animal Linguistics: A Primer. https://doi.org/10.31234/osf.io/ezmxq

Cartmill, E. A., & Byrne, R. W. (2010). Semantics of primate gestures: intentional meanings of orangutan gestures. *Animal cognition*, *13*(6), 793-804.

Fishbein, A. R., Fritz, J. B., Idsardi William J. & Wilkinson Gerald S. 2020 What can animal communication teach us about human language? *Phil. Trans. R. Soc. B, 375*, 20190042

Fitch, W. T. (2020). Animal cognition and the evolution of human language: why we cannot focus solely on communication. *Phil. Trans. R. Soc. B 375,* 20190046

Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., & Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLoS biology*, *16*(2), e2004825.

Graham, K. E., Furuichi, T., & Byrne, R. W. (2020). Context, not sequence order, affects the meaning of bonobo (*Pan paniscus*) gestures. *Gesture*, *19*(2-3), 335-364.

Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology*, *24*(14), 1596-1600.

Seyfarth, R., & Cheney, D. (2018). Pragmatic flexibility in primate vocal production. *Current Opinion in Behavioral Sciences, 21*, 56-61.

Smith, W. J. (1965). Message, meaning and context in ethology. *The American Naturalist, 99*, 405-409.

# ICONIC WORDS MAY BE COMMON IN EARLY CHILD INTERACTIONS BECAUSE THEY ARE MORE ENGAGING

KIRSTY GREEN*[1] and MARCUS PERLMAN*[2]

*Kirsty Green: kirsty.green@warwick.ac.uk
*Marcus Perlman: m.perlman@bham.ac.uk
[1]Department of Psychology, University of Warwick, Coventry, UK
[2]Department of English Language and Linguistics, University of Birmingham, Birmingham, UK

Spoken vocabularies contain a substantial number of iconic words, and a complete account of language evolution must explain how and why this is the case. Iconicity is especially prevalent in early communication, with one hypothesis being that this is because it scaffolds word learning by helping infants to establish referentiality, and another being that the phonological simplicity of iconic words makes them more pronounceable for infants. This study develops methods to examine whether another function of iconicity could be that it increases infants' engagement in interactions because there is something inherently fun about iconic communication. Iconicity ratings were assigned to transcribed words in a mother's utterances in 41 sessions with her infant between the ages of 12 and 24 months (from the Providence corpus (Demuth et al., 2006; MacWhinney 2000)). High and low iconicity bouts were identified by calculating the rolling average of iconicity per 5 words and focusing on those utterances with the highest and lowest averages. Key features from 58 high and 58 low iconicity bouts were then coded from corresponding video. Results showed that high iconicity interactions contained more behaviors considered to be indicative of engagement: eye contact, additional paralinguistic features, joint attention, smiling/laughing, gesture, child-directed utterances, and less displacement. These preliminary findings speak to the multi-functional nature of iconicity in parent-child interactions. They suggest that iconic words might be prevalent in current spoken vocabularies because people, especially young language learners and their caregivers, find these words to be intrinsically fun and engaging.

## 1. Background

Despite the longstanding doctrine of the 'arbitrariness of the sign', the vocabularies of spoken (as well as signed and protactile) languages are now understood to be characterized by iconicity, as well as arbitrariness (Dingemanse et al. 2015; Perniss et al. 2010; see Edwards 2017 for protactile). That is, the forms of many words are perceived to bear some amount of resemblance to what they mean, such as words like *woof* and *beep* that convey aspects of sound, and also words in other semantic domains like *wiggle* expressing manner of movement, and *teeny* expressing small size. Such words are especially prevalent in the speech of young children and their caregivers (Laing, 2020; Perry et al. 2018). In light of accumulating evidence of iconicity in spoken vocabularies, it is clear that a complete account of language evolution must explain, not just how iconicity could have grounded the first spoken words, but also why current vocabularies continue to be iconic, and why some kinds of words are more iconic than others.

One hypothesis is that iconicity is maintained in languages because it helps young children learn their first words by "bootstrapping" the connection between the form of a word and its referent (Imai & Kita, 2014; Perniss & Vigliocco 2014). If languages evolve to be more learnable by children (Christiansen & Chater, 2008), then iconic form-meaning pairings that are more learnable may be preferred in the lexicon over time. There is considerable evidence for the bootstrapping hypothesis. Children acquire iconic forms early and both they and their caregivers use them disproportionately often (Motamedi et al., 2020; Perlman et al., 2017; Perry et al., 2018). And critically, experiments show that iconicity can help children learn new words (e.g., Miyazaki et al., 2013). For example, 3-year-old English speaking children better learned Japanese verbs when they were iconic (Kantarzis et al 2011).

Another reason that iconic words – specifically, sound-imitative words or *onomatopoeia* – might be especially prevalent in early child vocabularies is because they tend to be phonological simple and easy to produce (Laing, 2014). An analysis of diary and video data from six languages showed that infants' early lexicons were dominated by a set of easily pronounceable prosodic structures – most often consonant harmony and consonant vowel – and that onomatopoeia fit these structures more often than non-onomatopoeic forms (Laing, 2019).

Here, we explore an alternative to these two hypotheses for why early child communication is so iconic. We consider whether iconic words might occur frequently in child communication because they are especially fun and engaging (*whee!*). More than prosaic words, iconic words tend to activate the senses and emotions (Nielsen & Dingemanse, 2020). Japanese speakers exposed to mimetics perceive meanings more vividly, describing experiencing an 'at-the-scene' feeling when hearing or reading these words, as though they contain, in some sense, a degree of 'sensory input' (Kita, 1997, p. 381). Moreover, Dingemanse and Thompson (2020) found that words rated high for iconicity also tended to be rated high for 'funniness'. They suggest that iconic words are 'valued as much for their performative character as their informative content' (p. 205).

Thus, iconic words could play a role in early word learning because they are more engaging for infants and their caregivers. There may be something inherently more fun about iconic communication that could enrich language-learning and facilitate the acquisition of early words. We explored this possibility by comparing infant-caregiver interactions with iconic words to those with arbitrary words, asking whether high iconicity interactions show more engagement by child and mother. To find out, we employed a novel method of using lexical iconicity ratings to continuously track the level of iconicity throughout an interaction, identifying bouts of high/low iconicity, and then coding these for different indicators of engagement by child and caregiver.

## 2. Methods

The data were taken from the Providence corpus (Demuth et al., 2006), part of the CHILDES database (MacWhinney 2000), a corpus of spontaneous video-recorded and transcribed interactions from six pairs of American English-speaking parents and children in their homes. As a proof of concept, we focused on a sub-corpus of 41 sessions between the infant 'Naima' and her mother (203201 total words). The recordings were made fortnightly while Naima was between the ages of 0;11.26 – soon after the onset of her first words – and 1;11.26.

From the mother's speech (155004 words), we identified bouts of high and low iconicity by using the mean iconicity ratings of her words used in the interactions (see Figure 1). These were taken from Winter, et al. (2022), which collected ratings for over 14,000 English words by asking adult native speakers to rate words on a scale of 1-7, with 1 being 'not iconic at all' and 7 being 'very iconic'. The average iconicity rating of each word was used to calculate a running average of iconicity per five words across the interactions. Iconicity ratings were assigned to the lemmatized forms of all words in the mother's utterances. Any words for which no iconicity rating existed were excluded (10.3% of words in the corpus). We set a threshold of a running average of 5 or above for bouts of high iconicity and 2 or below for bouts of low iconicity, which resulted in 58 bouts of each (high iconicity bouts = 2586 words; low iconicity bouts = 2676 words).

We next looked at the video of these instances to identify the features that characterized these different types of interactions. To assess the level of engagement, we coded these bouts for: eye contact (mother and child looking directly at each other during an utterance), joint attention (mother and child both directing their gaze towards the same object during an utterance), smiles and/or laughter, gestures and/or actions performed on an object by the mother and by the child, paralinguistic features (distinct moderations to tone, pitch, emphasis or extra-phonetic features such as panting like a dog), displacement (mother mentioning referents not present in the immediate context during an utterance) and child-directed speech (directly addressed to the child, as opposed to another adult, the camera or herself).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 3.7 | 4.06 | 3.62 | 3.7 | | | |
| | 3.3 | 3.8 | 2.1 | 5.7 | | | |
| **MOT:** | savoring | apples | and | blueberries | | | |
| | 4.12 | | | | | | |
| | 5.7 | NA | | | | | |
| **MOT:** | oh | mm | | | | | |
| | 3.944 | 4.1688 | 4.86 | 4.3 | 3.75 | 3.6 | 3.88 |
| | 4 | 3.4 | 5.5 | 2.9 | 4.3 | 1.9 | 4.8 |
| **MOT:** | have | another | bite | it | is | so | good |
| | 3.58 | | 4.1 | | | | |
| | 4 | NA | 5.5 | | | | |
| **MOT:** | have | a | bite | | | | |
| | 4.75 | 5.3 | 5.64 | 6.14 | | | |
| | 5.7 | 6.5 | 6.5 | 6.5 | | | |
| **MOT:** | oh | **yum** | **yum** | **yum** | | | |
| | 6.18 | 5.08 | | | | | |
| | 5.7 | 6.3 | | | | | |
| **MOT:** | **oh** | **yummy!** | | | | | |

Figure 1: An excerpt of transcribed speech from Naima's mother, showing the calculation of a rolling average of iconicity. The bottom number in red shows the average iconicity of the word below it, and the top number in blue shows the rolling average of iconicity. The high iconicity portion of the utterance is in bold.

## 3. Results

Table 1 summarizes the coded features observed within the high and low iconicity bouts. High iconicity bouts showed greater evidence of engagement than low iconicity bouts across all features. High iconicity interactions were characterized by more eye contact between mother and child, more joint attention on objects, more smiles and laughs by the child, and more gestures and actions on objects by the child and by the mother. In addition, the mother's speech contained more paralinguistic features, was more focused on the 'here-and-now' (less displacement), and was almost always child-directed.

Table 1. Features of utterances of high and low iconicity.

| | *High Iconicity* | *Low Iconicity* |
|---|---|---|
| Eye contact | 38% | 12% |
| Joint attention on an object | 45% | 36% |
| Number of smiles/ laughs by CHI | 16 | 6 |
| Number of gestures/ actions on objects by CHI | 29 | 10 |
| Number of gestures/ actions on objects by MOT | 47 | 35 |
| Number of interesting paralinguistic features by MOT | 49 | 9 |
| MOT's utterances that contain displacement | 4% | 38% |
| MOT's utterances that are child-directed | 99% | 92% |

## 4. Discussion

Iconic words are now known to be prevalent across spoken vocabularies, being especially common in interactions between young children and caregivers. One reason for this may be that iconic words are easier for children to learn, with iconicity providing an intuitive link between form and meaning. Additionally, onomatopoeic words in particular might be especially common because they are adapted to be some of the easiest words for young children to produce. However, it may also be that iconic words occur more frequently in this context because they are fun and more engaging for young children and their caregivers.

To explore this possibility, we developed a novel approach to investigate the role of iconicity in early child interactions, using iconicity ratings of words to identify high and low iconicity bouts during the natural flow of discourse. The findings showed consistently that high iconicity interactions were indeed more engaging. Compared to low iconicity interactions, high iconicity interactions typically involved caregiver and child as more active participants, more focused on each other and their shared environment, and openly enjoying the communicative situation. These results raise questions for future research into engagement as a function of iconicity and how it figures into early vocabulary development.

One notable finding – pointing to a question for future study – was that high iconicity interactions tended to be focused on the here-and-now, rather than discussing referents and topics outside of the present context. This finding differs from Perniss et al. (2017), who found British Sign Language-signing caregivers modified iconic signs more in non-ostensive contexts than ostensive contexts. While these contrasting results could reflect a difference between the modalities of speaking and signing, they might also be explained by other differences between the studies. Perniss et al. used an experimental design where they asked caregivers to imagine talking to their child about certain toys, both with the toys present and absent. They therefore controlled the topic and ensured that caregivers' signing would be specifically directed to the (imagined) child. In comparison, in our more naturalistic data, not all the mother's utterances were directed at her child – for example, sometimes another adult was present, or the mother spoke introspectively. This suggests that iconicity's role can be nuanced depending on context. While, as in Perniss et al.'s (2017) study of BSL, iconic modification may help a child identify a referent in a non-ostensive context, iconicity can serve other functions in ostensive contexts, such as increasing a child's enjoyment of a communication that focuses on the here-and-now.

Our proof-of-concept study has only analyzed a set of interactions between a single (English-speaking) parent and child. Future studies – expanding not just to more infant-caregiver dyads, but to interactions from widely diverse languages and cultures – are needed to understand the functions of iconicity in early language learning and development. However, even in this narrow demonstration, we think there is value to the novel method tested here, using word ratings to track the running level of iconicity over the course of conversation. This opens a new approach to studying iconicity in infant and caregiver interactions, and in many other discourses, that operationalizes iconicity in a more continuous fashion. Such an approach views iconicity, not just in terms of individual words and gestures, but as a broader quality of discourse in which the level of iconicity fluctuates over time.

## 5. Conclusion

We have suggested that iconic words may be especially common in early child interactions because they create an overall more engaging experience that motivates infants to communicate more. To the extent this is true, it implies that a different sort of selection pressure – not just learnability and ease of articulation – may operate on iconic words to maintain them in spoken lexicons. These words might continue to populate spoken languages because people, especially young language learners and their caregivers, find it to be intrinsically fun and engaging to use iconicity. Perhaps then the emergence of spoken language was partly rooted in an evolving sense of enjoyment that our hominin ancestors derived from the use of iconic vocalizations and what eventually became the first iconic words.

## References

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31,* 489-509.

Demuth, K., Culbertson, J., and Alter, J., (2006). Word-minimality, Epenthesis, and Coda Licensing in the Acquisition of English. *Language & Speech*. 49. pp. 137-174.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Science, 19,* 603-615.

Dingemanse, M. and Thompson, B., (2020). Playful iconicity: structural markedness underlies the relation between funniness and iconicity. *Language and Cognition*. 12(1). pp. 203-224. doi:10.1017/langcog.2019.49.

Edwards, T. (2017). Sign-creation in the Seattle DeafBlind community: A triumphant story about the regeneration of obviousness. *Gesture*, *16*(2), 305–328.

Imai, M. and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B*. 369. doi: http://dx.doi.org/10.1098/rstb.2013.0298.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound- symbolism facilitates word learning in English-speaking children. *Cognitive Science, 35,* 575–586.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*. 35. pp. 379-415.

Laing, C.E. (2014). A phonological analysis of onomatopoeia in early word production. *First Language, 34*(5), 387-405.

Laing, C.E. (2019). A role for onomatopoeia in early language: evidence from phonological development. *Language and Cognition, 11,* 173-187.

MacWhinney, B. (2000). The CHILDES project: tools for analyzing talk (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates

Motamedi, Y., Murgiano, M., Perniss, P., Wonnacott, E., Marshall, C., Goldin-Meadow, S. and Vigliocco, G. (2020). Linking language to sensory experience: Onomatopoeia in early language development. *Developmental Science*.

Miyazaki, M., Hidaka, S., Imai, M., Yeung, H.H., Kantartzis, K., Okada, H., and Kita, S. (2013). The facilitatory role of sound symbolism in infant word learning. in M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.). *CogSci 2013: the 35th Annual Conference of the Cognitive Science Society* (pp. 3080–3085). Cognitive Science Society.

Nielsen, A. and Dingemanse, M (2020). Iconicity in Word Learning and Beyond: A Critical Review. *Language and Speech*. pp.1–21.

Perlman, M., Fusaroli, R., Fein, D., & Naigles, L. (2017). The Use of Iconic Words in Early Child-Parent Interactions. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 6.

Perniss P, Thompson T, Vigliocco G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol*. 1, 1–15. (doi:10.3389/fpsyg.2010.00227).

Perniss, P., Lu, J. C., Morgan, G., and Vigliocco, G. (2017). Mapping language to the world: The role of iconicity in the sign language input. *Developmental Science*.· 21. e12551.

Perniss, P. and Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Phil. Trans. R. Soc. B*. 369: 20130300.

Perry, L.K., Perlman, M., Winter, B., Massaro, D.W., Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*. 2018-05. Vol.21(3).

Winter, B., Lupyan, G., Perry, L.K., Dingemanse, M., and Perlman, M. *Iconicity ratings for 14,000+ English words*. OSF. https://osf.io/y3wtx/

# SELECTIVE FORCES IN LANGUAGE EVOLUTION

Juan A. Guerrero Montero*[1], Andres Karjus[2], Kenny Smith[1], and Richard A. Blythe[1]

*Corresponding Author: j.a.guerrero-montero@sms.ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, United Kingdom
[2]School of Humanities, Tallinn University, Tallinn, Estonia

The quantitative study of language evolution and change has seen great development, with models like Iterated Bayesian Learning (Griffiths & Kalish, 2007) and the Utterance Selection Model (Baxter et al., 2006) admitting analytical and computational explorations of fundamental questions in the field (Kirby, 2000; Thompson et al., 2016; Nowak et al., 2001). With the advent of massive digital historical corpora, data-driven methods add a new empirical perspective on the fundamental forces that shape human language over long timescales.

A key question in evolutionary linguistics is how competition between linguistic variants shapes linguistic structure. The Wright-Fisher model from population genetics provides a simple but powerful description of competition, accounting for both stochasticity in inter-generational transmission and selective forces (Crow & Kimura, 1970). This model is justified for its use in linguistic scenarios through its derivation from Iterated Bayesian Learning (Reali & Griffiths, 2010).

Nevertheless, the Wright-Fisher model is not easy to apply to corpus data since approximations are necessary for computational efficiency. These typically assume a normal distribution, which is inaccurate in certain scenarios (Karjus et al., 2020). Here, we introduce a robust algorithm based on the Beta-with-Spikes (BwS) approximation to transition probabilities in the Wright-Fisher model, whose analytical form is given by Tataru et al. (2015). In figure 1.A, we illustrate the nuance and robustness of this method by comparing to other algorithms in the analysis of a reference data set, namely competition between past forms of verbs from the Corpus of Historical American English (COHA).

We further apply the BwS method to two data sets extracted from the Google Books corpus (Michel et al., 2011). We first quantify the effect of competing motivations on language structure. We analyse the competition between inflectional simplicity (i.e. regularity) and phonological simplicity in the past forms of English verbs whose irregular forms coalesce to avoid repeating of /d/ or /t/ sounds (e.g. he knit instead of he knitted). The BwS algorithm quantifies the net force arising from these competing motivations through a selection strength parameter, revealing that selection towards phonologically simpler irregular forms is more

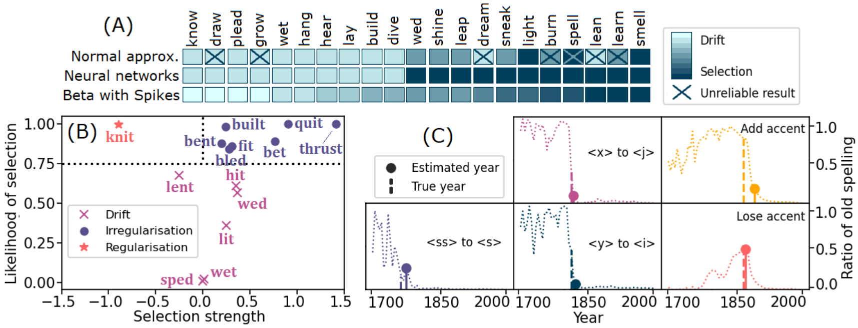


Figure 1. (A) Performance of BwS algorithm with those utilising normal approximations (Feder et al., 2014) and neural networks (Karsdorp et al., 2020), with reference to the likelihood of selection in competition between regular and irregular past forms of COHA verbs. BwS aligns with the robust neural network results, while providing a nuanced continuous classification instead of a binary one. Normal approximation is often unreliable due to non-normally distributed empirical frequency increments. (B) Application of BwS to competition between past tense verb forms that undergo coalescence in their irregular form. Positive selection strength implies favouring the irregular, phonologically simpler form, while the opposite is true for negative selection strength. Selection is concluded when likelihood is above 75%. Phonological simplicity beats inflectional simplicity in most cases, with half of all verbs in the data set showing high likelihood of selection towards irregularity, while only one shows regularisation. (C) Application of BwS to Spanish spelling reforms in the Google Books corpus. All reforms are detected with an error of 25 years or less, with three of them having an error of 10 years or less, even when trajectories are noisy (Real Academia Española, 1763, 1815, 1870).

common. If replicated in other domains, this result may shed light on how selection at different levels of linguistic structure might play out in language evolution.

Language is not only shaped by cognitive and internal factors, but also by social pressures that may change over time. This is illustrated in our second application, the dynamics of Spanish word spelling before and after reforms introduced by the Real Academia Española. Here, a new socially motivated bias appears that heavily favours a new spelling. This is reflected in the model by a change in the selection strength, thereby quantifying the level of acceptance of the reform by the literate population. Via likelihood maximisation, the BwS algorithm detects the year of introduction of the spelling reform with high precision (see figure 1.C). Thus this algorithm can detect and quantify major changes in social dynamics in the data, even when the exact date or origin of those changes is unclear.

In summary, the BwS algorithm applies to both socially and internally motivated competition, quantifies the net selective force arising from competing motivations and detects variations in this force over time. It provides a tool for the numerical analysis of diachronic linguistic data that is more insightful and reliable than previous methods. The continued development of such numerical tools opens the door to the empirical analysis of the social, cognitive and internal pressures that shape the structure of language.

## References

Baxter, G. J., Blythe, R. A., Croft, W., & McKane, A. J. (2006). Utterance selection model of language change. *Phys. Rev. E*, *73*, 046118.

Crow, J., & Kimura, M. (1970). *An introduction in population genetics theory.* New York: Harper and Row.

Feder, A., Kryazhimskiy, S., & Plotkin, J. (2014). Identifying signatures of selection in genetic time series. *Genetics*, *196*(2), 509–522.

Griffiths, T., & Kalish, M. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, *31*, 441–480.

Karjus, A., Blythe, R., Kirby, S., & Smith, K. (2020). Challenges in detecting evolutionary forces in language change using diachronic corpora. *Glossa*, *5*.

Karsdorp, F., Manjavacas, E., Fonteyn, L., & Kestemont, M. (2020). Classifying evolutionary forces in language change using neural networks. *Evolutionary Human Sciences*, *2*.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight (Ed.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 303–323). Cambridge University Press.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Nowak, M., Komarova, N., & Niyogi, P. (2001). Evolution of universal grammar. *Science*, *291*(5501), 114–118.

Real Academia Española. (1763). *Ortografía de la lengua castellana* (3 ed.). Madrid, Spain.

Real Academia Española. (1815). *Ortografía de la lengua castellana* (8 ed.). Madrid, Spain.

Real Academia Española. (1870). *Prontuario de ortografía de la lengua castellana en preguntas y respuestas* (1 ed.). Madrid, Spain.

Reali, F., & Griffiths, T. L. (2010). Words as alleles: connecting language evolution with bayesian learners to models of genetic drift. *Proc. R. Soc. B*, *277*, 429-–436.

Tataru, P., Bataillon, T., & Hobolth, A. (2015). Inference under a wright-fisher model using an accurate beta approximation. *Genetics*, *201*, 1133–1151.

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*, 4530–4535.

# UNDERSTANDING DESIGN FEATURES OF MUSIC AND LANGUAGE: THE CHORIC/DIALOGIC DISTINCTION

FELIX HAIDUK[1], and W. TECUMSEH FITCH[1,2]

[*]Corresponding Author: felix.haiduk@univie.ac.at
[1]Department of Behavioral and Cognitive Biology, University of Vienna, Vienna, Austria
[2]Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria

Music and spoken language share certain characteristics: both consist of sequences of acoustic elements that are combinatorically combined, and these elements partition the same continuous acoustic dimensions (frequency, formant space and duration). However, the resulting categories differ sharply: scale tones and note durations of small integer ratios appear in music, while speech uses phonemes, lexical tone, and non-isochronous durations. Why did music and language evolve to become the two systems we have today, differing in these specific features? Based on a reverse-engineering perspective on design features of language and music proposed by Hockett (1960) and Fitch (2006), we developed a framework suggesting that design features of prototypical music and language are a response to their differential deployment along three different continuous dimensions, which we call the goal dimension, the novelty dimension and the interactivity dimension. Also, we suggest that the hybrid nature of intermediate systems like poetry, chant, or solo lament follows from their deployment in the less typical context. At the core of all three dimensions are predictive processes about linguistic or musical sequences unfolding in time (see Koelsch et al., 2019; Kuperberg & Jaeger, 2016). Predictions involve uncertainty and surprise/ information which can be quantified using information theory (Shannon, 1948). Trajectories of information and uncertainty at multiple levels of the unfolding sequence thus form the basis of our theorizing.

(1) The goal dimension concerns the broader purpose of linguistic or musical sequence productions, whether to convey semantic messages, or to elicit and modulate aesthetic responses in a broad sense (see Huron, 2016). Both language and music can be deployed in propositional and aesthetic contexts, and similar responses follow: with more propositional goals, the multiple levels of the speech or musical sequence are more interdependent, and vary their information density to support successful inference of propositional content. For aesthetic goals, independent variation across levels enables more unconstrained variation in uncertainty and surprise, effectively exploiting the human reward system.

(2) The novelty-repetition dimension involves the repeatability or novelty of (groups of) elements and their relations at different scales and at multiple levels of musical or linguistic sequences. While language usually allocates repeatability to the phonological level and novelty to the morphosyntactic and semantic levels (related to propositionality), music typically allows both novelty and repeatability across all levels of the musical sequence, and meter seems to be especially crucial as a predictive layer throughout, enabling both prediction and surprise.

(3) We argue that the familiar goal and novelty dimensions alone are not sufficient to explain differences in design features between music and language: the interactivity between individuals is crucial. The interactivity dimension, the poles of which we term 'choric' and 'dialogic', concerns the temporal coordination of linguistic or musical productions of multiple participants. For dialogic contexts the only coordinative constraint concerns the timing of the turn-taking between individuals. Thus, we predict a lower information density and thus higher predictability towards the end of phrases, across all levels of the sonic stream, for both music and language. This aligns well with neural markers like the N400 (e.g. Grisoni et al., 2017), changes in speech rate at turn completion points in conversations (Walker, 2010; Wightman et al., 1992) or the notion of musical closure in harmony and melody (Huron, 2006). In contrast, choric performance requires tight temporal coordination of all contributing individuals, enabled by high predictability in timing and frequency of sonic events (Keller et al., 2014). Isochronous meter and discrete pitches in scales are design solutions that enable a group of participants to join in making a coherent sound sequence, allowing both novelty and repeatability. Simultaneous speaking of multiple talkers requires uttering of precisely the same words at the same time, as happens for example in religious chanting. Attention should then be much more focussed on coordination than in dialogic speech acts, an isochronic and/or metrical scaffolding should develop (cf. Bowling et al., 2013), and body motion, facial expressions or prosodic intonation should be more pronounced in a spoken choric context.

Our framework avoids an overly simplistic dichotomy between language and music by also encompassing non-canonical forms of music and language like chant, poetry, or exchange of musical solos. It supports comparisons of different forms of communication across distinct modalities and can help to generate new hypotheses about optimal design of signals satisfying multiple different requirements. We hope that it will also be fruitfully employed in animal communication research, especially for species engaging in chorusing or duetting, such as various bird species, dolphins, bats, or gibbons, broadening the scope of comparisons with music and/or language.

Bowling, D. L., Herbst, C. T., & Fitch, W. T. (2013). Social origins of rhythm? Synchrony and temporal regularity in human vocalization. *PLoS ONE, 8*(11), e80402. https://doi.org/10.1371/journal.pone.0080402

Fitch, W. T. (2006). The biology and evolution of music : A comparative perspective. *Cognition, 100*, 173–215. https://doi.org/10.1016/j.cognition.2005.11.009

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. *The Journal of Neuroscience, 37*(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Huron, D. (2006). *Sweet Anticipation*. Cambridge, Massachusetts: MIT Press.

Huron, D. (2016). Aesthetics. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (2 ed., pp. 233–245). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.19

Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130394. https://doi.org/10.1098/rstb.2013.0394

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences*, *23*(1), 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Walker, G. (2010). The phonetic constitution of a turn-holding practice: rush-throughs in English talk-in-interaction. In D. Barth-Weingarten, E. Reber, & M. Selting (Eds.), *Prosody in Interaction. Studies in discourse and grammar (23).* (pp. 51–72). Amsterdam, Philadelphia: John Benjamins Publishing Company. https://doi.org/https://doi.org/10.1075/sidag.23

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717.

# PRESSURE FOR AND AGAINST NON-ARBITRARINESS: EVIDENCE FROM PHONESTHEMES

DAVID A. HASLETT*[1], ZHENGUANG G. CAI[1,2]

*Corresponding Author: david.a.haslett@gmail.com
[1]Dept. of Linguistics and Modern Languages, The Chinese University of Hong Kong
[2]Brain and Mind Institute, The Chinese University of Hong Kong

Words with related meanings tend to sound dissimilar, and this arbitrariness allows for large vocabularies that make nuanced distinctions (like *glint* versus *shine*). Languages violate arbitrariness in constrained ways, such as patterns in form that provide cues indicating word class (e.g., English nouns and verbs tend to be stressed on different syllables; Kelly & Bock, 1988), but phonesthemes (i.e., sound clusters associated with semantic features, like /gl/ with light, as in *glint* and *gleam*) more directly defy arbitrariness by pertaining to word meaning proper. We provide evidence from two experiments that phonesthemes influence how people process word meanings (e.g., they're slower to decide that *glove* is unrelated to light, compared to a control like *mitten*) and that phonesthemes aid in the recognition of visually degraded words. These findings support that, in English, non-arbitrariness interferes with word meaning access and that the lexicon evolves to balance competing needs: Arbitrariness maintains subtle distinctions, but non-arbitrariness can facilitate recognition in noisy environments.

## 1. Introduction

Word forms tend to be arbitrarily assigned to meanings (de Saussure, 1983; Hockett, 1960). For example, *bench* and *sofa* refer to similar concepts but sound nothing alike. Evidence from simulations and novel word learning suggests that this arbitrariness occurs not by chance but in response to communicative pressure. When people need to distinguish referents, cultural evolution increases arbitrariness (Kirby, Tamariz, Cornish, & Smith, 2015), which is consistent with the finding that arbitrariness makes it easier to acquire a large vocabulary (Gasser, 2004) and to learn specific word meanings (Monaghan, Christiansen, & Fitneva, 2011). There is selective pressure for an arbitrary lexicon. But non-arbitrariness can be advantageous, too, by facilitating categorization. For example, people read nouns in sentences faster if those nouns sound more like other nouns than like verbs (Farmer, Christiansen, & Monaghan, 2006). Importantly, such patterning in form (i.e., systematicity) pertains to low-level semantics, like word class (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015), and so is functionally compatible with arbitrariness at the level of individual word meanings, like *bench* and *sofa* (Monaghan et al., 2011).

However, words that overlap substantially in meaning (not just that are members of the same broad category) sometimes sound similar, as in phonesthemic words. Phonesthemes are sound clusters associated with semantic features, like /gl/ with light (e.g., *glimmer* and *glaze*; Firth, 1930), and there is speculation that words which contain phonesthemes but lack the corresponding features, such as *gland* and *glove* (unrelated to light), are apt to acquire that feature (Bolinger, 1965; Blust, 2003). For example, English words beginning with the phonestheme /fl/ (as in *flap* or *flow*) have grown more likely to relate to "moving through the air" (Smith, 2016). Such semantic changes defy the tendency for the lexicon to evolve to be more arbitrary or to limit non-arbitrariness to low-level semantics. Novel word learning experiments are consistent with this corpus analysis. People tend to pair novel words containing phonesthemes (e.g., *glep*, like *glow* and *glimmer*) with definitions containing semantic features associated with that phonestheme (e.g., something related to light), and when asked to invent labels for such definitions, people are more likely to produce novel words containing the corresponding phonesthemes (Hutchins, 1998; Abelin, 1999; Magnus, 2001).

Phonesthemes might motivate semantic change in an efficient lexicon even if the resulting non-arbitrariness hinders our ability to distinguish similar concepts (like *glint* versus *gleam*). Consider how ambiguity is sometimes advantageous: Languages reuse sound sequences that are easier to produce, and while this results in homophones, it eases the burden of articulation on speakers (Piantadosi, Tily, & Gibson, 2012). Similar-sounding words with related meanings might facilitate comprehension, too. For example, Bergen (2004) found that phonesthemic primes sped up recognition of targets containing the same phonestheme (*glow* → *glitter*) more than formal (*glove*) or semantic (*shine*) primes did.

There is suggestive evidence that phonesthemes motivate semantic change and facilitate word recognition, but those studies rely on correlations and single phonesthemes (Smith, 2016; Blust, 2003), on novel words rather than known words learned in rich, realistic contexts (Hutchins, 1998; Magnus, 2001; Abelin, 1999), and on priming rather than the ambient influence of the structure of the lexicon (Bergen, 2004). And while previous studies have found that non-arbitrariness facilitates processing of part of speech (e.g., Farmer et al., 2006), those experiments do not apply to words that overlap in high-level conceptual semantics. We therefore conducted two experiments to provide evidence that phonesthemes affect how people process the meanings of English words and that the resulting non-arbitrariness facilitates recognition in some conditions. All data and scripts are available at osf.io/sna92.

## 2. Experiment 1

We first investigated whether people associate a word containing a phonestheme (e.g., /gl/) with the corresponding semantic feature (light) even when the word is unrelated to that feature (*glove*). In an online experiment (implemented on Gorilla.sc; Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2018), 57 adult native English speakers (recruited from Prolific.co) decided whether a cue word associated with a phonestheme's feature (e.g., *shine*, with light) was semantically related to a target that either contains that phonestheme (*glove*) or is a control (*mitten*). We manipulated the target (pseudo-phonesthemic vs. control), for a one-factor, two-level design, manipulated within items and within subjects. In each trial, the cue (e.g., *shine*) was unrelated to the target (*glove*/*mitten*), and the pseudo-phonesthemic and the control targets were synonyms matched on semantic distance from the cue, as measured by a distributional semantic model (computed using the LSAfun package in R; Günther, Dudschig, & Kaup, 2015). The targets were also matched on formal similarity to cues (Levenshtein distance), word frequency, age of acquisition, concreteness rating, bigram frequency, orthographic neighbourhood density (OLD20), and length (all from the English Lexicon Project; Balota et al., 2007). We divided 126 items (unrelated cue–target pairs) and 126 fillers (related pairs) into six lists, each comprising 21 pseudo-phonesthemic targets, 21 control targets, and 42 fillers. We assigned each participant to one list. To incentivize fast and accurate responses, participants had only two seconds to make each decision, with three seconds of feedback following timeouts and incorrect responses. We excluded participants with accuracy below 75% (seven of 64 participants) and trials faster than 200ms (one trial). Linear mixed effects modelling conducted using the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2017) indicates that pseudo-phonesthemic targets were judged to be unrelated to cues significantly more slowly than controls were (see Tables 1 and 2 and Figure 1). Importantly, semantic relatedness decisions involve word meaning access, not just recognition or low-level categorization, so these effects stem from a phenomenon that is distinct from statistical regularities within a broad category, as Dingemanse et al. (2015) define systematicity. Phonesthemes pertain to word meaning proper and so are incompatible with arbitrariness.

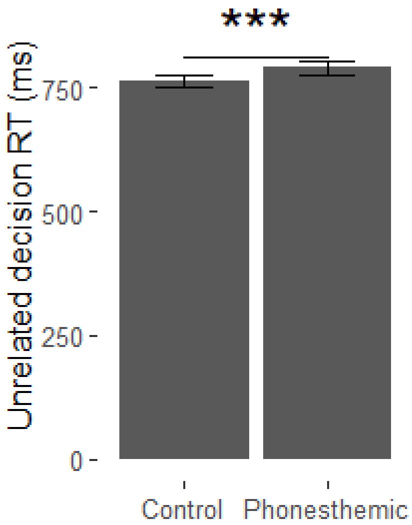


Figure 1: Semantic relatedness decision RT (correct unrelated decisions only) for targets containing phonesthemes (*glove*) versus arbitrary controls (*mitten*) following unrelated cues (*shine*) in Experiment. 1. Error bars indicate 95% confidence intervals.

Table 1: Comparison of RT by target type (sum contrast coded: pseudo-phonesthemic = 1, control = -1) for correct unrelated semantic relatedness decisions in Experiment 1, fitting the random intercepts of subject and item, with target type added as a random slope to the random intercept of subject, as justified by forward model comparison (cf. Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

| | *β* | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | 775.18 | 16.21 | 47.8 | < .001 |
| Pseudo-phonesthemic | 13.59 | 3.79 | 3.6 | < .001 |

Table 2: Comparison of accuracy (proportion of unrelated decisions) by target type (sum contrast coded: pseudo-phonesthemic = 1, control = -1) for semantic relatedness decisions in Experiment 1, fitting the random intercepts of subject and item, with target type added as a random slope to both random intercepts, as justified by forward model comparison.

| | *β* | *SE* | *z* | *p* |
|---|---|---|---|---|
| Intercept | 2.95 | 0.17 | 17.3 | < .001 |
| Pseudo-phonesthemic | -0.16 | 0.15 | 1.1 | .284 |

## 3. Experiment 2

We next investigated whether phonesthemes facilitate word recognition in noisy conditions. In another online experiment, 59 participants made lexical decisions about phonesthemic words (e.g., *glitter*, related to light and containing the /gl/ phonestheme) or controls (*sparkle*) that were presented clearly or were visually degraded (i.e., the first and second half of the word flickering separately on the screen, each half seen twice for 100ms). It was a 2 (target: phonesthemic vs.

control) x 2 (presentation: clear vs. degraded) design, with both factors manipulated within subjects and presentation manipulated within items. Target type was manipulated between items, but phonesthemic targets were paired with synonymous control targets and, as in Experiment 1, were matched on word frequency, AoA, concreteness, OLD20, bigram frequency, and length. We evenly divided 216 items (108 phonesthemic and 108 control targets) and 216 non-word fillers into eight lists and assigned each participant to one list. We again incentivized fast, accurate responses with a time limit and penalties, with the same exclusion criteria as Experiment 1, excluding five of 64 participants. Logistic and linear mixed effects modelling indicates that decisions were significantly faster and more accurate for phonesthemic than control words and were faster for degraded words that contained phonesthemes (i.e., a significant interaction of target type and presentation; see Tables 3 and 4 and Figure 2). Note that while the main effect of target type is significant in both analyses, the simple effect of target type (i.e., when using treatment coding, with control targets and clear presentation as the reference levels, thereby analyzing the effect of target type in the clear condition; cf. Brehm & Alday, 2022) is not significant in the RT analysis, which suggests that the advantage for phonesthemes in the degraded condition is the real source of the main effect on RT (see supplementary scripts on osf.io/sna92).

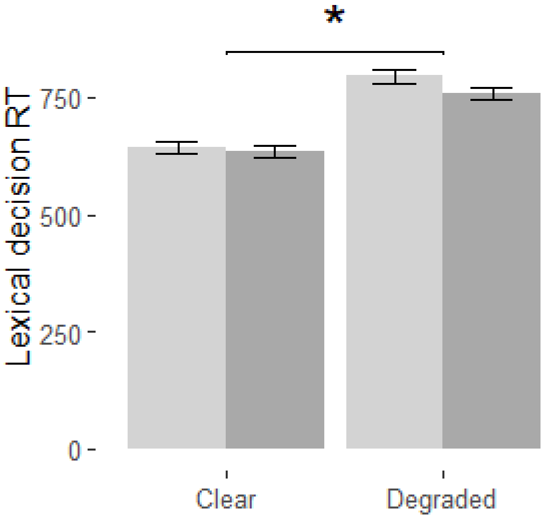


Figure 2: Lexical decision RT (correct decisions only) for phonesthemic targets (dark grey; e.g., *glitter*) versus controls (light grey; e.g., *sparkle*) presented clearly or visually degraded in Experiment 2. Error bars indicate 95% confidence intervals.

Table 3: Comparison of RT by target type (sum contrast coded: phonesthemic = 1, control = -1) and presentation type (visually degraded = 1, clear = -1) for correct lexical decisions in Experiment 2, with presentation type added as a random slope to the random intercepts of subject and item, as justified by forward model comparison.

| | *β* | *SE* | *t* | *p* |
|---|---|---|---|---|
| Intercept | 710.14 | 14.24 | 49.9 | < .001 |
| Degraded presentation | 69.36 | 4.03 | 17.2 | < .001 |
| Phonesthemic target | -11.52 | 4.18 | 2.8 | .006 |
| Degraded : Phonesthemic | -6.71 | 2.90 | 2.3 | .022 |

Table 4: Comparison of accuracy by target type (sum contrast coded: phonesthemic = 1, control = -1) and presentation type (visually degraded = 1, clear = -1) in Experiment 2, fitting the random intercepts of subject and item, with presentation type added as a random slope to the random intercept of item, as justified by forward model comparison.

| | *β* | *SE* | *z* | *p* |
|---|---|---|---|---|
| Intercept | 2.95 | 0.17 | 16.9 | < .001 |
| Degraded presentation | -0.47 | 0.12 | 4.0 | < .001 |
| Phonesthemic target | 0.44 | 0.13 | 3.3 | .001 |
| Degraded : Phonesthemic | -0.04 | 0.07 | 0.5 | .583 |

## 4. Discussion and Conclusion

We found that phonesthemes affect how people process word meanings. In Experiment 1, participants were slower to decide that words containing phonesthemes (e.g., *glove*, compared to *mitten*) were unrelated to meanings typically associated with that phonestheme (e.g., *shine*). We also found that phonesthemes confer an advantage in noisy conditions. In Experiment 2, people made faster lexical decisions about visually degraded words when those words were phonesthemic than when they were controls (e.g., *glitter* versus *sparkle*).

Experiment 1 supports that non-arbitrariness can interfere with word meaning access (e.g., Gasser, 2004; Monaghan et al., 2011) and is consistent with studies suggesting that phonesthemes, and similar-sounding words in general, motivate semantic change (Smith, 2016; Blust, 2003; Haslett & Cai, 2021). As form influences interpretation, subtle shifts in word meanings can accumulate. At first glance, these two implications seem to be at odds. If non-arbitrariness impedes processing, then it should decrease rather than increase over generations, assuming the lexicon evolves for the sake of efficiency (e.g., Gibson et al., 2019). However, Experiment 2 can help resolve this apparent inconsistency. While non-arbitrariness slows access to the precise meaning of a word, it facilitates recognition in challenging conditions. Compare this to the polysemy advantage in lexical decisions. People are faster to recognize words with multiple related senses (e.g., *plant* means both “situate” and “organism with roots”) than homonyms with multiple unrelated meanings (e.g., *bark* in the context of a dog versus a tree), perhaps because related senses share a broad attractor basin in semantic space (Rodd, Gaskell, & Marslen-Wilson, 2002, 2004). We found that words which sound similar (rather than identical) to words with related meanings confer the same sort of advantage when words are harder to read. It is possible, though entirely speculative, that words' overlapping in both form and meaning is reflected in the organization of semantic space, providing a conglomerate target when encountering those words in noisy conditions.

These are small effects (a 28ms disadvantage for pseudo-phonesthemic words in Experiment 1, a 35ms advantage for phonesthemic words in the degraded condition in Experiment 2), but they are on par with comparable studies. For example, Rodd et al. (2002) reported a 29ms disadvantage for homonyms, compared to unambiguous words, and a 33ms advantage for polysemous words with many senses, compared to those with few senses. Are differences on the order of tens of milliseconds sufficient to drive language evolution, leading people to favour some words over others or to alter words' sounds or meanings? Such minute advantages, even if imperceptible to the mind's naked eye, are consistent with what Christiansen and Chater (2016) call the now-or-never bottleneck: We need to process words fast in order to shuttle information along. Still, how fast is fast enough is an open empirical question, one that must weigh the value of word recognition against distinguishing word meanings.

Previous research has explored the competing pressures of word meaning individuation and word learning (e.g., Monaghan et al., 2011; Kirby et al., 2015). Here, we have provided evidence that patterns in sound and meaning influence interpretation and can thereby impede word meaning access but that those patterns also facilitate recognition in noisy conditions. There are communicative pressures both for and against phonesthemes, depending on depth of processing.

## References

Abelin, Å. (1999). Phonesthemes in Swedish. In *Proceedings of XIV International Conference of Phonetic Sciences*, *99,* 1333-1336.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods, 52*(1), 388-407.

Balota, D., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445-459.

Bergen, B. (2004) The Psychological Reality of Phonesthemes. *Language, 80*(2), 291-311.

Blust, R. (2003). The Phonestheme ŋ- in Austronesian Languages. *Oceanic Linguistics, 42*(1), 187-212.

Bolinger, D. (1965). *Forms of English: Accent, Morpheme, Order*. Harvard UP.

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language,* 125.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences, 39*.

Farmer, T., Christiansen, M., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*(32), 12203-122208.

Firth, J. R. (1930). *Speech*. Oxford UP.

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the Annual Conference of the Cognitive Science Society, 26*.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun – An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods, 47,* 930-944.

Haslett, D. & Cai, Z. (2021). New neighbours make bad fences: Form-based semantic shifts in word learning. *Psychonomic Bulletin & Review*.

Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of experimental psychology: human perception and performance, 14*(3), 389.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87-102.

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26.

Magnus, M. (2001.) *What's in a word? Evidence for phonosemantics*. Dissertation, University of Trondheim.

Monaghan, P., Christiansen, M. H., & Fitneva, S. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General, 140*(3), 325-347.

Piantadosi, S., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*(3), 280-291.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of memory and language, 46*(2), 245-266.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive science, 28*(1).

de Saussure, F. (1983). *Course in General Linguistics*. Open Court.

Smith, C. (2016). Tracking semantic change in fl- monomorphemes in the *Oxford English Dictionary*. *Journal of Historical Linguistics*, *6*(2), 165-200.

# THE STUDY OF ONOMATOPOEIA FROM THE COMPARATIVE DEVELOPMENTAL PERSPECTIVE

HAYASHI MISATO*[1,2], YAMADA TAKEMI[1], UMEDA YUSUKE[1], TANABE KEIKO[3], TAKAMURA MAKI[3], and NAKAJIMA KENSUKE[3]

*Corresponding Author: hayashi-misato@chubu-gu.ac.jp
[1]Faculty of Education, Chubu Gakuin University (CGU), Kakamigaraha, Gifu, Japan
[2]Japan Monkey Centre, Inuyama, Aichi, Japan
[3]Hokuriku Gakuin University, Kanazawa, Ishikawa, Japan

## 1. Introduction

Onomatopoeia is the use of words that imitate the sounds associated with the objects or actions they refer to. Onomatopoeias have been attracting attentions in various fields in human linguistics (e.g., Laing, 2019), but have not previously been explored from the cognitive developmental aspects for species comparisons. This study examines the origin and context of onomatopoeia and determines its significance in linguistic development from the evolutionary viewpoint.

### *1.1. A review on existing literatures on onomatopoeia*

The Japanese language has a wide range of onomatopoeic expressions. Various onomatopoeias have been used to convey subtle information that cannot be expressed in normal words. Among the five categories used in a review on onomatopoeias in Japanese articles (Yamada et al., 2022), we specifically focused on the following two areas. (ii) Psychology and human relations: onomatopoeias have been referred to as sound symbolism (Imai & Kita, 2014), facilitating linkage between a linguistic sound to a perceptual image in early childhood. (iii) Physical expression: onomatopoeias can be easily combined with physical/bodily actions and gestures, thus, facilitate sharing vivid images.

## 2. Methods and results of observational study on human children

We conducted observations and video recordings in two settings by focusing on onomatopoeias in the social interactions involving children in Japanese culture.

### *2.1. Onomatopoeias in mother-child interactions*

Observations on mother-child interactions took place in a facility providing a playroom and toys to mothers and children, located in CGU. We focused on an episode which included onomatopoeic and gestural expressions from a three-year-old child during playing house. Some gestural patterns were linked to onomatopoeias, indicating the emergence of rudimentary forms of linguistic categorization on internal action images in early childhood.

### *2.2. Onomatopoeias in interactions among peers*

Observations on social interactions among peers took place in a kindergarten. We focused on an episode which included onomatopoeic expressions from four-year-old children recorded during play activities. Children's vocalizations may express their feelings in communicative contexts as action words of onomatopoeias.

## 3. Implications for the comparative studies on cognitive development

Japanese children used a rich repertoire of onomatopoeias to communicate about their internal images accompanied by gestures or bodily actions. Onomatopoeias may bridge the babbling or simple actions in infancy with symbolic word utterances, iconic gestures, or symbolic plays in early childhood, thus, have a scaffolding effect on the linguistic development. Since onomatopoeias connect simple vocalizations/gestures to the subsequent linguistic abilities in humans, more observations in early childhood will highlight the evolutional continuity with gestural communications or manipulative plays in nonhuman primates.

## References

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B,* 369, 20130298.

Laing, C. (2019). A role for onomatopoeia in early language: evidence from phonological development. *Language and Cognition*, 11, 173-187.

Yamada, T., Hayashi, M., Yageta, K., Umeda, Y., Mizuno, Y., Dalrymple, N., Nakajima, K., Tanabe, K., & Takamura, M. (2022) The study of onomatopoeia from the developmental perspective: The process of language production. *The Journal of Chubu Gakuin University and Chubu Gakuin College,* 23, 77-89.

# EXPLORING OUR INNER-APE: SHARED UNDERSTANDING OF GESTURE BETWEEN CHIMPANZEES AND HUMANS

MATTHEW G. HENDERSON*[1], CATHERINE HOBAITER[1], KIRSTY E. GRAHAM[1], PATRICK GEORG GROSZ[2], and PRITTY PATEL-GROSZ[2]

*Corresponding Author: mh295@st-andrews.ac.uk
[1]School of Psychology and Neuroscience, University of St Andrews,
St Andrews, Scotland

[2]Department of Linguistics and Scandinavian Studies, University of Oslo,

Oslo, Norway

## 1. Introduction

Decades of great ape gesture research revealed various parallels to human language, such as diverse repertoires of intentional signals used flexibly across varied, everyday interactions (Pika et al, 2005; Liebal, Pika, & Tomasello, 2006; Hobaiter & Byrne, 2011). Evidence for a shared origin to ‘ape’ gestural communication is supported by extensive repertoire overlap between extant apes, which increases between more phylogenetically related species (Byrne & Cochet, 2011; Graham et al, 2018). While it remains unclear how human fits within this shared system of gesture, borrowing methods from ape gesture research seems promising. For example, Kersken and colleagues (2019) used ‘ape-typical’ observational methods to find that children aged 1-2 years share 89% of their natural gesture use with chimpanzees (46/52 gestures).

To understand how human development and ability of language affects our retention of innate ‘ape’ gestures, we use a novel online experiment that tests human understanding of chimpanzee gestures when they are performed by other humans in a controlled environment. Using human models allows us to control additional cues and contexts present in the ape environment that may affect participant performance, while also ensuring that participants do not accommodate their interpretation of the signaler’s intended gestural meaning based on personal perceptions and knowledge of nonhuman great apes.

## 2. Methods

We pilot a novel online experiment (hosted on Gorilla.sc) that tests how adult humans interpret the intended meaning of 10 established chimpanzee gestures (Hobaiter et al, 2014; Byrne et al, 2017) when performed by expert researchers. Each video had participants choose one of four potentials 'meanings', with the correct response(s) varying depending on the gesture's natural flexibility. There were four *Tight gestures* with one correct response, and six *Flexible gestures* with two correct responses (see Table 1). Trial order was randomized but each gesture's available responses were consistent. Binomial tests were used to determine if participants were above chance at detecting correct chimpanzee response(s) – note that we used the following corrected values of hypothesized chance based on the available correct responses: *Tight gestures*=0.25, *Flexible gestures*=0.5; and *overall* (all gestures combined)=0.4.

## 3. Results

We analyzed n=3000 responses from n=300 participants. Overall, participants performed above chance across all 10 gestures (chance=0.4; binomial n=0.568, p<0.001, LL=0.553, UL=1.000) but performed comparatively better on the six *Flexible gesture*s (chance=0.5, binomial: n=0.740, p<0.001, LL=0.722, UL=1.000) than the four *Tight gesture*s (chance=0.25, binomial: n=0.309, p<0.001, LL=0.287, UL=1.000). Binomial tests for each gesture revealed significance for 6/10 gestures (see Figure 1); 5/6 *Flexible gestures* (arm swing, beckon, grab hold, hit object, reach palm) and 1/4 *Tight gestures* (hand fling); with four gestures at or below chance (arm raise, hand shake, push, reach wrist).

## 4. Discussion

Our experiment shows successful use of a new 'proof of concept' approach in understanding the root of 'ape' gesture, and supports the hypothesis that humans retain some access to this shared system. Higher adherence of chimpanzee responses in *Flexible gestures* than *Tight gestures* suggests our evolution may have favored retaining more ambiguous gesture-types, and the mixed results across individual gestures may represent our specified range of overlap to chimpanzees. However, additional factors may influence our results, such as our specific selection of gestures (we avoided unusual gestural forms and meanings for this type of typical human interaction, such as 'pirouette' and 'sexual-solicitation'). Additionally, *Tight gestures* naturally have greater specificity that may not have fit well within the context we presented (two adult humans resting on a couch); while the less-specified *Flexible gestures* could include natural use within similar contexts. Overall, we demonstrate practical adaptation of studying chimpanzee gestural retention in humans and our results may hint at a shared, underlying 'ape' gestural origin to human language.

## 5. References

Byrne, R. W., & Cochet, H. (2017). Where have all the (ape) gestures gone? Psychonomic Bulletin & Review, 24, 68–71.

Byrne, R. W., E. Cartmill, Genty, E., Graham, K. E., Hobaiter, C., Tanner, J.E. (2017). Great ape gestures. Simple intentional communication with a rich set of innate signals. Animal Cognition, 20, 755-769.

Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., & Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. PLOS Biology, 16(2), e2004825.

Hobaiter, C., & Byrne, R. W. (2011). The gestural repertoire of the wild chimpanzee. Animal Cognition, 14(5), 745–767.

Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. Current Biology, 24, 1596–1600.

Kersken, V., Gómez, J. C., Liszkowski, U., Soldati, A., & Hobaiter, C. (2019). A gestural repertoire of 1-to 2-year-old human children: in search of the ape gestures. Animal Cognition, 22(4), 577-595.

Liebal, K., Pika, S., & Tomasello, M. (2006). Gestural communication of orangutans (Pongo pygmaeus). Gesture, 6(1), 1-38.

Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). Gestural communication of apes. Gesture, 5(1-2), 41-56.

## 6. Acknowledgements

We would like to thank everyone who helped contribute to make this research possible: the members of the Wild Minds Lab and the Super Linguistics Group; everyone who selflessly shared and participated in this experiment; and the nonhuman apes whose study allowed for our insight into their gestural systems. This research received funding from the European Union’s 8th Framework Programme, Horizon 2020, under grant agreement no 802719.

# HORIZONTAL TRANSFER OF LINGUISTIC FEATURES AS AN INTEGRAL PART OF THE EVOLUTIONARY HISTORY OF LANGUAGE FAMILIES – PHYLOGENETIC RECONSTRUCTION OF LINGUISTIC TREE–BASED NETWORKS

Lina Herbst*[1,2,3], Volker Gast[1], Denise Kühnert[2,3], and Mary Walworth[2,3]

*Corresponding Author: lina.herbst@uni-jena.de
[1]Department of English and American Studies, Friedrich Schiller University Jena, Jena, Germany
[2]Max Planck Institute for the Science of Human History, Jena, Germany
[3]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

More than 7000 different languages from more than 200 language families are spoken worldwide today. The evolutionary relationships within a language family are often illustrated and studied using a linguistic phylogeny, also called a language family tree. Based on linguistic features this allows inference of the internal structure and the time-depth of a language family. However, linguistic features are not always “inherited”, they can also be acquired horizontally through language contact.

Non-treelike evolution is also common in evolutionary biology. Scholars have represented these horizontal relationships within phylogenetic networks, a tree with horizontal edges. By allowing linguistic features to move horizontally, a more accurate representation of linguistic evolution is possible. However, in historical linguistics network reconstructions are still in their infancy and no standard has been established so far.

Pioneering work by Nelson-Sathi et al. uncovered hidden lexical borrowing in Indo-European (IE) languages, showing that borrowing is more widespread than previously thought (Nelson-Sathi et al., 2011). A software library based on this approach was provided by List and Moran (List & Moran, 2013). In addition, minimal lateral networks (MLNs) were applied to 40 IE languages by investigating presence-absence patterns of cognate sets (List, Nelson-Sathi, Geisler, & Martin, 2014). The MLN approach uses weighted parsimony and thus allows for a certain proportion of parallel evolution (List et al., 2014). List et al. have also investigated the history of 40 Chinese dialects using the MLN approach (List, Nelson-Sathi, Martin, & Geisler, 2014). In addition, Willems et al. used distance data from IE languages to infer hybridization networks (Willems et al., 2016). Cathcart et al. introduced a novel method for investigating areal dispersal of linguistic features, and they applied this approach to 117 IE languages (Cathcart, Carling, Larsson, Johansson, & Round, 2018). In 2019, a mixture-model approach was applied to the Austronesian, Sinitic, IE and Japonic language families to detect non-tree-based signals (Verkerk, 2019), and there are even more examples for the detection of linguistic transfer events and the reconstruction of linguistic networks.

In this work, we develop a novel process that generates rooted and timed linguistic networks in which horizontal edges account for transfer events between languages. These networks will be based on a birth-death process, which will be adapted to account for horizontal transfers to detect the timing and the location of linguistic transfer. The novel process will be integrated in an open-source Bayesian phylogenetic framework, such that knowledge about the past can be included e.g. on the depth of a sub-family. We will evaluate and quantify the effects horizontal transfer has on different aspects of language from the grammar to the lexicon. The process will be applied to the Polynesian language family, which has been suggested to have a highly reticulate history with many linguistic transfers. This will shed light on the timing and phylogenetic placement of linguistic transfers among the Polynesian languages and reveal which linguistic features were transferred.

## Acknowledgements

Lina Herbst was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 461279234.

## References

Cathcart, C., Carling, G., Larsson, F., Johansson, N., & Round, E. (2018). Areal pressure in grammatical evolution. *Diachronica. International Journal for Historical Linguistics*, *35*(1), 1–34.

List, J.-M., & Moran, S. (2013). An open source toolkit for quantitative historical linguistics. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 13–18.

List, J.-M., Nelson-Sathi, S., Geisler, H., & Martin, W. (2014). Networks of lexical borrowing and lateral gene transfer in language and genome evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *36*(2), 141–150.

List, J.-M., Nelson-Sathi, S., Martin, W., & Geisler, H. (2014). Using phylogenetic networks to model chinese dialect history. *Language Dynamics and Change*, *4*(2), 222–252.

Nelson-Sathi, S., List, J.-M., Geisler, H., Fangerau, H., Gray, R. D., Martin, W., & Dagan, T. (2011). Networks uncover hidden lexical borrowing in indo-european language evolution. *Proceedings. Biological Sciences*, *278*(1712), 1794–1803.

Verkerk, A. (2019). Detecting non–tree–like signal using multiple tree topologies. *Journal of Historical Linguistics*, *9*(1), 9–69.

Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Sciullo, A. M. D., & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. *BMC Evolutionary Biology*, *16*(180).

# SEMANTIC SIMILARITY SHAPES HOW AGREEMENT MARKERS SPLIT OVER POSITIONS IN VERB MORPHOLOGY

Borja Herce[1,2 † *], Carmen Saldana[1,2 †], and Balthasar Bickel[1,2]

[*] Corresponding Author: borja.hercecalleja@uzh.ch
[†] These authors have contributed equally to this work
[1] Department of Comparative Language Science, University of Zurich, Switzerland
[2] Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

Agreement markers that refer to the same feature or argument tend to be found in the same position within inflectional forms (e.g., all subject agreement in suffixes, all object agreement in prefixes; Mansfield, Stoll, & Bickel, 2020); however, little is known about the exceptions to this trend. In this study, we explore the positional properties of subject and object person-number agreement markers in a phylogenetically diverse sample of 227 languages from 97 different stocks—based on AUTOTYP (Bickel et al., 2017) plus additional data collected from the WALS 100-language sample (Dryer & Haspelmath, 2013). The result is 325 person-number paradigms, whose agreement morpheme order was surveyed to explore trends and split patterns. We find that a majority of agreement paradigms only require reference to a single position, thus obeying the principle of CATEGORY CLUSTERING (Mansfield et al., 2020). A sizeable minority (128 paradigms, about 39%), however, show what we call POSITIONAL SPLITS, whereby different person-number bundles are marked in different positions (e.g., prefix or suffix) within the paradigm (as illustrated in Table 1). We ask whether positional splits deviate from category clustering in systematic ways and whether their recurrence is proportional to the amount of shared feature values.

Table 1. Different patterns of syncretism in person-number verbal paradigms.

| | NATURAL PATTERN | | L-TYPE PATTERN | | X-TYPE PATTERN | |
|---|---|---|---|---|---|---|
| | GUMER, *open* IPFV | | KOASATI, *hear* ACT | | BASQUE, *walk* PRS | |
| | SG | PL | SG | PL | SG | PL |
| 1 | **ə-**kəft | **nɨ-**kəft**-ɨnə** | há:lo**-l** | **il-**há:l | **na-**bil | **ga-**bil**-tza** |
| 2 | **tɨ-**kəft | **tɨ-**kəft**-o** | **is-**há:l | **has-**há:l | **za-**bil**-tza** | **za-**bil**-tza-te** |
| 3 | **yɨ-**kəft | **tɨ-**kəft**-o** | há:l | há:l | **da-**bil | **da-**bil**-tza** |

Our survey suggests three patterns: natural, L-type and X-type patterns, which are illustrated with orange, blue and green cells in the examples from Gumer (Ethiopia), Koasati (US) and Basque (Spain) in Table 1. In natural patterns, all forms with agreement affixes in the same position(s) share at least one feature value throughout, e.g., prefixal forms share SG and circumfixal forms share PL in

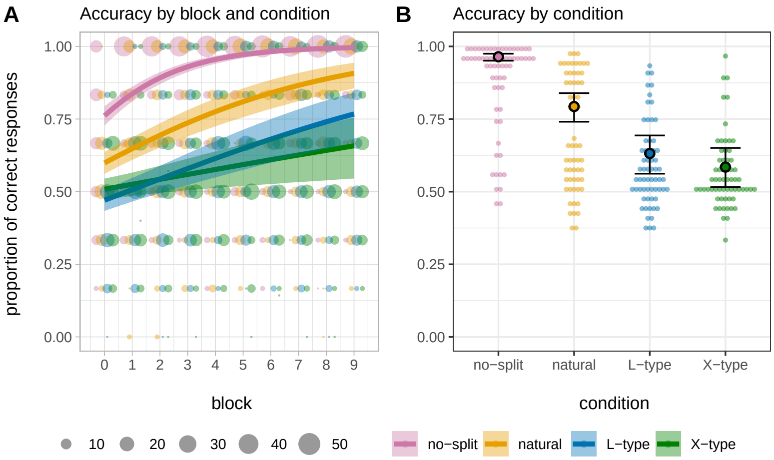


Figure 1. Raw accuracy scores (colour faded) and Bayesian mixed model mean estimates (solid). Error bars and shaded areas show the 90% credible intervals of the mean estimates.

Gumer. In L-type patterns, forms with the same positional properties share values only partially, e.g., two of the prefixal forms in Koasati share PL but differ in person; two share 2, but differ in number. X-type patterns share even less feature values, e.g., circumfixal forms in Basque contain one form that shares no feature value with the other forms. We fitted Bayesian mixed-effects models comparing the occurrences of each type (natural, L-type and X-type) to the occurrences we would expect by chance from all logically possible configurations in person-number 3×2 paradigms (with language and family as random slopes). We find that the natural patterns are over-represented in our data, the most unnatural X-patterns are underrepresented, and intermediate-naturalness L-patterns occur with a similar frequency as expected by chance.

These observations suggest a bias towards natural over L over X patterns when languages evolve over time and space. We hypothesize that this bias is grounded in a preference for semantic similarity during the transmission and learning of word forms. To test this hypothesis we conducted an online artificial language learning experiment (N=247) where we trained and tested participants on a person-number verbal agreement paradigm with positional splits according to natural vs. L-type vs. X-type patterns (Saldana, Herce, & Bickel, 2022, January 21). We ran a further control condition with no splits (i.e., following CATEGORY CLUSTERING). Results are consistent with the hypothesized learnability gradient *no-split* > *natural* > *L-type* > *X-type* (see Figure 1), thus matching the observed cross-linguistic tendencies. Our findings support the notion that semantic similarity shapes the evolution and transmission of morphological structure (Dautriche, Mahowald, Gibson, & Piantadosi, 2017; Mansfield et al., 2020) and that it does so in a gradient way.

## References

Bickel, B., Nichols, J., Zakharko, T., Witzlack-Makarevich, A., Hildebrandt, K., Rießler, M., Bierkandt, L., Zúñiga, F., & Lowe, J. B. (2017). *The AUTOTYP typological databases. Version 0.1. 0.* Retrieved from https://github.com/autotyp/autotyp-data/tree/0.1.0.

Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science*, *41*(8), 2149–2169. https://doi.org/10.1111/cogs.12453.

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *Wals online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from https://wals.info/.

Mansfield, J., Stoll, S., & Bickel, B. (2020). Category clustering: A probabilistic bias in the morphology of verbal agreement marking. *Language*, *96*(2), 255–293. https://doi.org/10.1111/cogs.13107.

Saldana, C., Herce, B., & Bickel, B. (2022, January 21). *Learnability of morphological patterns of syntagmatic splits.* [Preregistration] https://doi.org/10.17605/OSF.IO/YZCXP.

# BEYOND BROCA: NEURAL ARCHITECTURE AND EVOLUTION OF A DUAL MOTOR SPEECH COORDINATION SYSTEM

GREGORY HICKOK [*1,2], JONATHAN VENEZIA 2[2,3], and ALEX TEGHIPCO 3[3]

[*]Corresponding Author: greg.hickok@uci.edu
[1]Departments of Cognitive Sciences & Language Science, University of California, Irvine, CA USA
[2] VA Loma Linda Health Care System and Dept. of Otolaryngology, Loma Linda University School of Medicine, Loma Linda, CA, USA
[3]Department of Communication Sciences and Disorders, University of South Carolina, Columbia, SC USA

## 1. Introduction and preview

Classical neural architecture models of speech production propose a single system coordinating all the vocal articulators from lips to larynx. Here we propose a dual coordination system in which laryngeal control of pitch-related aspects of prosody and song are controlled by a dorsal precentral gyrus system while supralaryngeal articulation at the phonetic/syllabic level is controlled by a ventral precentral system. The separability of these neural systems supports models of the evolution of speech and language in which song-based communication was an early foundational stage.

.

## 2.0 Evidence for distinct two systems

Evidence for the existence of two speech coordination systems comes from functional imaging (Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Price et al., 1996; Wilson, Saygin, Sereno, & Iacoboni, 2004), direct cortical stimulation mapping (Lu et al., 2021), and post stroke apraxia of speech (Graff-Radford et al., 2014; Hickok et al., 2014), all of which have identified speech-related regions in both the ventral and more dorsal central gyrus. We refer to these as the dorsal and ventral precentral speech areas (dPCSA and vPCSA).

## 3.0 Evidence for a voice pitch-related coordination in the dPCSA

Two recent published findings argue for a functional distinction between the two speech areas, specifically that the dPCSA, but not vPCSA is involved in voice pitch coordination for prosody and song. One is an intracranial recording study showing that neural activity in the dPCSA (but not vPCSA) is correlated with control of voice pitch and when stimulated elicits vowel-like vocalization (Dichter, Breshears, Leonard, & Chang, 2018). The other is an fMRI study showing that the dPCSA (but not vPCSA) has *auditory* spectrotemporal receptive fields that code pitch-related information (Venezia, Richards, & Hickok, 2021). We also provide new functional connectivity evidence: dPCSA has greater connectivity to auditory cortex and vPCSA has greater connectivity to secondary somatosensory cortex.

## 4.0 The Dual Speech Coordination Model

The observations described above led us to the following dual coordination model:

<u>The dPCSA</u> is part of a sensorimotor control circuit for pitch-related vocalization, which would include prosodic aspects of speech as well as song, and predominantly acts on the laryngeal effector via the dorsal laryngeal motor cortex.
<u>The vPCSA</u> is part of a sensorimotor control circuit for syllabic and phonetic-related speech features and acts on supralaryngeal effectors as well as voicing-related control of the larynx via the ventral laryngeal motor cortex.

We further propose and provide evidence for the following secondary hypotheses.

-Both the dPCSA and vPCSA are part of hierarchy of speech control networks in frontal cortex with both situated between lower-level primary motor systems posteriorly and higher-level systems anteriorly, extending into the posterior middle frontal gyrus and inferior frontal gyrus (Broca’s area), respectively.
-The location of the dPCSA can be explained in evolutionary terms, having evolved out of a voluntarily controllable goal-directed orienting hub in premotor cortex, which is present in both human and non-human primates, and which makes use of pitch and other acoustic cues.
-The vPCSA evolved out of voluntarily controlled supralaryngeal effectors, used extensively for mastication, oral-grasping behaviors, and non-vocal oral communication (e.g., lip smacking) with the major advance for speech being the coordination of open-close jaw cycles (the syllabic frame) with articulatory gestures (the phonetic content), as proposed in MacNeilage’s frame/content model speech production evolution (MacNeilage, 1998).
-The dPCSA and its associated circuit for controlling vocal pitch, prosody, and song evolved prior to the vPCSA system’s control over phonetic articulatory gestures and represented an early stage of the evolution of language, as originally proposed by Darwin (Darwin, 1871).

## References

PREPRINT: https://psyarxiv.com/tewna/

Hauser, M. D., & Fitch, W. T. (2003). What are the uniquely human components of the language faculty? In M. H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 158-181). Oxford: Oxford University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707-784.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.

Dichter, B. K., Breshears, J. D., Leonard, M. K., & Chang, E. F. (2018). The control of vocal pitch in human laryngeal motor cortex. *Cell, 174*(1), 21-31 e29. doi:10.1016/j.cell.2018.05.016

Graff-Radford, J., Jones, D. T., Strand, E. A., Rabinstein, A. A., Duffy, J. R., & Josephs, K. A. (2014). The neuroanatomy of pure apraxia of speech in stroke. *Brain Lang, 129*, 43-46. doi:10.1016/j.bandl.2014.01.004

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience, 15*, 673-682.

Hickok, G., Rogalsky, C., Chen, R., Herskovits, E. H., Townsley, S., & Hillis, A. E. (2014). Partially overlapping sensorimotor networks underlie speech praxis and verbal short-term memory: evidence from apraxia of speech following acute stroke. *Front Hum Neurosci, 8*, 649. doi:10.3389/fnhum.2014.00649

Lu, J., Zhao, Z., Zhang, J., Wu, B., Zhu, Y., Chang, E. F., . . . Berger, M. S. (2021). Functional maps of direct electrical stimulation-induced speech arrest and anomia: a multicentre retrospective study. *Brain*. doi:10.1093/brain/awab125

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav Brain Sci, 21*(4), 499-511; discussion 511-446. doi:10.1017/s0140525x98001265

Price, C. J., Wise, R. J. S., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., . . . Friston, K. J. (1996). Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain, 119*, 919-931.

Venezia, J. H., Richards, V. M., & Hickok, G. (2021). Speech-Driven Spectrotemporal Receptive Fields Beyond the Auditory Cortex. *Hear Res, 408*, 108307. doi:10.1016/j.heares.2021.108307

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat Neurosci, 7*, 701-702.

# INCREMENTAL EVOLUTION OF SYNTAX: FROM ‘BIGRAMS’ TO ‘BRANCHING’

Dieter Hillert[*1]

[*]Corresponding Author: dhillert@gmail.com
[1]School of Speech, Language, and Hearing Sciences,
San Diego State University, USA

Here, we explore how syntax might have evolved after the hominin lineage branched off from our closest living relatives, monkeys and nonhuman great apes. Our approach focuses on ‘bigrams’ and ‘branching’ in terms of unordered, external and internal Merge in association with workspace and memory storage resources. The syntactic abilities of monkeys and nonhumans great apes but also various typologies of human syntax in children and adults indicate that the language capacity incrementally involved in the hominin lineage. Indirect paleoanthropological records let us moreover predict that already *Homo erectus* was equipped with a language capacity which was premodern or even modern.

## 1. Introduction

All organisms are adapted to communicate with one another (or sometimes across species). Cultural transmissions, however, can be found only in a subset of life-forms, including fish, insects, birds, monkeys, and apes. The development and acquisition of these behavioral cultures are based on species-species genetic endowments. However, we should not ignore that modern humans share with other primates or mammals most neurobiology (e.g., neuronal morphology, neurotransmitter, subcortical structures) and the vocal tract of monkeys are considered as ‘speech-ready’. Thus, the neurobiological differences between nonhuman primates and modern humans can be considered as a ‘needle in a haystack’ (Fitch, 2011; Fitch et al., 2016).

Modern humans (or possibly their closest extinct ancestors), however, added another dimension by inventing a language system to share subjective experiences or to express thoughts. Already, Lenneberg (1967) recognized that this dimension is based on an innate human-specific capacity, and Chomsky (1968) described the human language capacity as an innate ‘universal grammar’ underlying all human languages. Speakers make ‘infinite use of finite means’, that is, a finite set of syntactic rules are used to create an infinite number of sentences or text structures, aka “super-sentences” (Watumull et al., 2013). Our closest living relatives, the nonhuman great apes, clearly display superior cognitive abilities compared to monkeys or other animals. We explore for this reason how this innate capacity evolved in context of the expansion and refinement of (sub-) cortical circuits after the hominin lineage branched off from monkeys and nonhuman great apes.

## 2. Bigrams

The strong 'minimalist program' of generative linguistics explicitly claims that syntactic recursion is an innate property of UG (Chomsky, 1995; Hauser et al., 2002; Berwick & Chomsky, 2016). This recursion concept is based on the operation 'unbounded Merge', that is, it recursively merges words to create bottom-up larger phrases in a binary fashion. The simplest form of Merge is to take two distinct elements of the lexicon such as (X, Y), and combine them to an unordered binary set unit {X, Y}. It is a binary unit without a syntactic head which determines the syntactic category of a phrase. Headless binary sets are still marked by an unspecified (neutral) label. However, the pre-Merge conditions involve the ability to form and recall discrete elements and create bigrams.

Bigrams can be found not only in our closest living relatives, monkeys and nonhuman great apes, but they also (re-) surface in varieties of modern spoken and sign languages. Bigrams are rooted in non-linguistic cognition across primate species and represent the basic linear layer of human syntax across different domains (e.g, Jackendoff, 2003; Miyagawa, & Clarke, 2019). In addition to long-term 'memory storage' (MS) for lexical elements, our account considers the increase of temporary 'workspace' (WS), which is required for syntactic operations (e.g., Rizzi, 2016). We do not exactly know which behavioral and/or environmental changes triggered the expansion of (sub-)cortical regions and circuits in the hominin lineage, but the increase of brain structures goes hand-in-hand with an increase of WS capacity required for more complex intrinsic structure-building processes.

### *2.1. Monkey Syntax*

Monkeys primarily communicate by means of vocalizations. Their call repertoire consists of a small set of discrete alarms calls. One or more calls can be combined with another call, either discrete or suffixed. Female Diana monkeys *(Cercopithecus diana)*, for example, emit four calls whereas call *A*, which refers to caller identity, is systematically combined with the calls *L, H, R* to function as a social context modifier (Candiotti et al., 2012). Male putty-nosed monkeys *(Cercopithecus nictitans)* produce the general alarm call *P* (pyow) and the specific eagle-related alarm call *H* (hack). Both calls combined refer to a non-predator situation, whereas *H* follows always *P*. The number of *Hs* and *Ps* in a sequence informs about the distance of group movement (Schlenker et al., 2016). Cotton-top tamarins' *(Saguinus oedipus)* long-call combinations are emitted by socially isolated individuals to get in touch with their conspecifics. These contact calls consist of one or more chirp-like syllables followed by two or more whistle-like syllables. Again, two call types are combined in a fixed order and the variation is the number of call type repetitions (Cleveland, & Snowdon, 1982). Black-fronted titi monkeys *(Callicebus nigrifrons)* produce *A*-calls for predators in the canopy and *B*-calls for predators on the ground. The combination of both call types informs conspecifics not only about predator type but also about the location of the threat and the number of continuous *B*-calls might indicate probabilistic rather than categorial meaning (Casar et al., 2013; Berthet et al., 2019). These examples show that the syntax of monkeys is restricted to the combination of maximal two elements in a fixed order and their WS = 1 (e.g., Pasternak, & Greenlee, 2005; Wakita, 2020).

### *2.2. Nonhuman Ape Syntax*

Old World monkeys and apes branched of about 30-25 million yrs. ago (mya). Thus, we expect to find qualitative differences between both clades about their abilities to structure their ‘language’. Indeed, great apes coordinate often facial expressions, postures, vocalizations, or gestures, and they outperform monkeys on many cognitive tasks in an enculturated environment. The photographic memory capacities of chimpanzees (*Pan troglodytes*) are impressive. *Ai* and her son *Ayumu* outperformed humans in tasks which required to recall number sequences (Matsuzawa, 1985). Since the cognition of great apes’ centers on visuospatial information and their communication is rather multimodal than relying on vocalization, they have been primarily trained on sign language or graphic symbols (lexigrams) to examine their language capacities.

*Kanzi*, the famous bonobo (*Pan paniscus*), acquired the meaning of lexigrams as an infant by observing his foster mother’s lessons. It is said that he learned about 348 of those symbols. *Kanzi* accompanied often graphic communication with an articulatory but distorted equivalent. He had the ability to communicate without difficulties EVENT, OBJECT combinations such as HIDE ICE or WATER HIDE. He also produced new combinations such as CAR TRAILER meaning he wanted to be driven to the trailer (Greenfield, & Savage-Rumbaugh, 1991). *Kanzi* acquired also some American Sign Language by watching a video of the signing gorilla *Koko* who learned thousand signs and understood about 2,000 spoken English words (Patterson, 1978).

*Chantek* was an orang-utan (*Pongo pygmaeus*) who mastered to learn approximately 150 signs and initiated communication to meet his needs. His vocabulary matches that of a very young child (Miles, 1993). He learned single concepts such as people names, places, food items, actions, animals, pronouns, locations, attributes, etc. and immediately began to combine his acquired signs (e.g., He signed COKE DRINK after drinking a coke, or PULL BACK while pulling a caregiver's hair through the fence.) He began to over- or underextend the meanings of his signs (e.g., He used the sign DOG for dogs, barking noise, birds, horses, noisy helicopter, etc). *Chantek* also invented new sign combinations (e.g., EYE DRINK for contact lens solution). He clearly imaged objects or actions (e.g., He pointed to objects when he was asked WHERE HAT or WHAT WANT?). His signing was spontaneous, non-repetitious, did merely imitate his caregivers' signing, actively initiated signing to meet his needs.

Nonhuman great apes acquire conceptual representations of events and objects. In an encultured environment, they can flexibly combine two elements to express needs or ideas. Trigrams, however, are less often used. The order of the elements is in principle free but seems to follow mostly pragmatic strategies. For example, in DAVE MISSING FINGER the topic is 'Dave'; however, in HIDE ICE or WATER HIDE, the order EVENT or OBJECT seems to be subjective or random. The concepts are ordered linear, and it is unclear whether "grouping" takes place. WS of great apes is 2 ± 1 (Read et al., 2022).

## 3. Branching

In briefly reviewing monkey and nonhuman ape language, it is apparent that multiple components contributed to the evolution of syntax in modern humans. The continuous increase of cognitive resources in terms of WS and MS in the lineage of great apes (including humans) may be the main factors for the ability to generate complex structures between signals (cf. Zuberbühler, 2020). Moreover, complex structure-building processes can be found at all linguistic levels but also across non-linguistic domains (e.g., Fujita, 2017; Asano et al., 2021). The uniqueness of human syntax may be therefore based on an underlying universal computational capacity which includes enhanced WS/MS resources.

### *3.1. Human Syntax*

Typically developing (human) children produce at around the age of 18 months two-word utterances such as WANT-JUICE or CAR-GO which lack inflections and function words (Bloom, 1993). Some accounts assume that this stage reflects WS limitations (Blake et al., 1994). At this point, human and encultured ape syntax are quite similar in production: (a) No significant differences were found for the type of bigrams such as AGENT-EVENT, EVENT-OBJECT, EVENT-LOCATIVE; (b) Auditory WS of 2 ± 1 matches between panins and 2-year-old toddlers (Patkowski, 2014).

Adult human language differs from nonhuman ape language in two fundamental aspects. The human child has the innate ability to acquire abstract semantic, syntactic and phonological entries independent of immediate perceptual experiences and goals, and over time it can build or tap into additional cognitive resources to process complex sentence and discourse structures. Cowan (2001) argues therefore that only 4 ± 1 items can be held within focus of attention. Complex sentence processing ability is not enabled by WS per se but by its strategic subcomponents. However, human languages tap into these resources to different degrees.

For instance: Independent of particular language pairs, adult L2 learners acquire first a linear strategy with bigrams before they learn more complex structures. Broca's aphasic patients who suffer from agrammatism produce only telegraphic-like speech without inflections and function words. Bigrams are also the basic layer of particular sign languages. For example, the first generation of deaf children of the *Idioma de Señas de Nicaragua* combined signs to bigrams such as AGENT-EVENT, and the next generation elaborated on these structures (Senghas et al., 2004).

About ten million people speak the Malayan dialect *Riau Indonesian* which functions as a lingua franca. This language virtually has no syntactic categories or inflections, and the word order is based on semantic principles (Gil, 2005). For example, *ayam* means CHICKEN and *makan* EAT. The bigrams *makan ayam* or *ayam makan* are open to the same readings and they are unmarked for numerus, tense, aspect, semantic role, etc. Situational context is however sufficient that the speakers getting the meaning across. Speakers of the indigenous, agglutinative Amazonas language *Pirahã* get by without rearranging canonical structures and clear evidence of hierarchical syntactic structures is missing (Everett, 2009). In general, the spectrum of typologies from analytical resp. isolating languages to polysynthetic languages varies strongly with respect to the use of syntactic hierarchies and recursion.

### *3.2. Homo erectus*

Paleoanthropological data indicate an incremental neurobiological process. External evidence from phylogenetic and ontogenetic stages, let us hypothesize milestones of syntactic evolution. There are many more intermediary steps which need to be spelled out. Monkey's initial alarm response will be presumably replaced by a modifying response since they cannot keep more than one unit in WS. In contrast, nonhuman great apes acquire concepts rooted in perception and combine them into bigrams. Their lexicon may have an upper limit since they do not make use of compositional structures.

We assume that the next milestone broadly corresponds to *Homo erectus* since significant neurobiological and behavioral changes can be associated with this species. Based on indirect evidence, we predict an intermediate stage between of structure-building processes between nonhuman great apes and *Homo sapiens*. The appearance of the *C* variant of the protein *SRGAP2* which involves neuronal migration and differentiation is also associated with the rise of *Homo erectus* about 2 mya (Hillert, 2015; Martins et al., 2018). The endocasts of *Homo erectus* share a pronounced Broca's cap, the bulge that appears at the level of the temporal pole in *Homo sapiens*. The specimens' brain sizes, moreover, range between 600-1,000 cc and had human-like prefrontal and temporoparietal regions. A recent endocranial morphology study of *early Homo* supports the view that frontoparietal areas did not evolve separately but in concert (Ponce de León et al., 2021). *We must, therefore, assume that the genetic endowment of a (pre-) modern language capacity preceded cultural-linguistic accumulations.* Acheulean tool industry coincides with the appearance of *Homo erectus* about 1.8 mya and Acheulean hand axes produced in Africa might fall in the category of art as their making reflects planning and abstract thinking. In addition, *late Homo erectus* produced non-functional, aesthetic engravings and decorations (Stout et al., 2015; Mcbrearty & Brooks, 2000). Accumulated cultural processes began about 1 mya in the merging species *late Homo erectus* and *pre-archaic* resp. *superarchaic Homo sapiens* (e.g., Bergström et al. 2021).

In sum, indirect neurobiological and cognitive-behavioral changes indicate that *Homo erectus* was equipped with a language capacity, whether it was premodern or even modern, we cannot tell (Dediu, & Levinson, 2013; Hillert, 2021). Possible WS differences can be considered as marginal. They might have used basic structures such as a linear grammar and headless binary sets (unordered Merge). Since direct evidence is lacking, we reserve asymmetric grouping (external Merge) and recursion (internal Merge) to the language of *Homo sapiens*.

## 4. Conclusions

In discussing the biology capacity of syntax and language in general, it is important to separate the genetic endowment of species and the extent to which a species makes use of its cognitive capacities. The evolution of syntax, from bigrams to branching, indicates that multiple factors played a crucial role, including WS and MS resources. In particular, inner speech (rehearsal) may have contributed to WS and representational resources required for complex sentence and discourse processing. Syntax incrementally emerged in the hominin lineage, and we must therefore assume that *Homo erectus* was already equipped with a premodern or even modern language capacity.

## Acknowledgements

Supported by FY2020 JSPS invitational long-term fellowship at Kyoto University.

## References

Asano, R., Boeckx, C., & Seifert, U. (2021). Hierarchical control as a shared neurocognitive mechanism for language and music. *Cognition, 216,* 104847.

Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E.M.L., & Skoglund, P. (2021). Origins of modern human ancestry. *Nature, 590,* 229-237.

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi monkeys combine alarm calls to create probabilistic meaning. *Science Advance, 5,* eaav3991.

Berwick, R.C., & Chomsky, N. (2016). Why only us; language and evolution. Cambridge, MA: MIT Press.

Blake, J., Austin, W., Cannon, M., Lisus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development, 17 (1),* 91-107.

Bloom L. (1993). The transition from infancy to language: Acquiring the power of expression. Cambridge University Press.

Candiotti, A., Zuberbühler, K., Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition, 15,* 327-339.

Cäsar, C., Zuberbühler, K., Young, R.J., & Byrne, R.W. (2013) Titi monkey call sequences vary with predator location and type. *Biology Letters, 9,* 20130535.

Chomsky, N. (1968). *Language and Mind*. New York: Harper and Row.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Pess.

Cleveland, J., & Snowdon, C.T. (1982). The complex vocal repertoire of the adult cotton-top tamarin (Saguinus oedipus oedipus). *Zeitschrift für Tierpsychologie, 58 (3),* 231-270.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral Brain Science, 24,* 87-185.

Dediu, D., & Levinson, S.C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology, 4,* 397.

Everett, D.L. (2009). Pirahã culture and grammar: A response to some criticisms. *Language, 85,* 405-442.

Fitch, W.T. (2011). The evolution of syntax: an exaptationist perspective. *Frontiers in Evolutionary Neuroscience, 3,* 9.

Fitch, W.T., de Boer, B., Mathur, N. & Ghazanfar, A.A. (2016). Monkey vocal tracts are speech-ready. *Science Advances, 2 (12).*

Fujita, K. (2017). On the parallel evolution of syntax and lexicon: A Merge-only view. *Journal of Neurolinguistics, 43 B,* 178-192.

Gil, D. (2005). Word order without syntactic categories: How Riau Indonesian does it. In: A. Carnie, H. Harley, and S.A. Dooley (Eds.), *Verb first: On the syntax of verb-initial languages* (pp. 243-263). Amsterdam: John Benjamins.

Greenfield, P.M., & Savage-Rumbaugh, E.S. (1991). Imitation, grammatical development, and the invention of protogrammar by an ape. In N.A. Krasnegor, D.M. Rumbaugh, R.L. Schiefelbusch, and M. Studdert-Kennedy (Eds.), *Biological behavioral determinants of language development* (pp. 235-257). Hillsdale, NJ: Erlbaum.

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298 (5598)*, 1569-1579.

Hillert D. G. (2015). On the evolving biology of language. *Frontiers in Psychology*, *6*, 1796.

Hillert, D. (2021). How did language evolve in the lineage of higher primates? *Lingua*, *263*, 103158.

Jackendoff, R. (2003). Précis of foundations of language: Brain, meaning, grammar, evolution. *Behavioral and Brain Sciences, 26*, 651-707.

Lenneberg, E.H. (1967). *Biological foundations of language*. New York: John Wiley and Sons.

Martins, P.T., Maties, M., & Boeckx, C. (2018). SRGAP2 and the gradual evolution of the modern human language faculty. *Journal of Language Evolution, 3(1),* 67-78.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature 315*, 57-59.

Mcbrearty, S., & Brooks, A.S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of Human Evolution 39 (5)*, 453-563.

Miles, L.W. (1993). Language and the orang-utan: The old 'person' of the forest. In P. Cavalieri, & P. Singer (Eds.) *The great ape project* (pp. 42-57). New York: St. Martin's Griffin.

Miyagawa, S., & Clarke, E. (2019) Systems underlying human and old world monkey communication: one, two, or infinite. *Frontiers Psychology, 10,* 1911.

Pasternak, T., & Greenlee, M. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience, 6,* 97-107.

Patkowski, M. (2014). Looking for structure: Is the two-word stage of language development in apes and human children the same or different? *Studies in Second Language Learning and Teaching, 4 (3)*, 507-528.

Patterson, F.G. (1978). The gestures of a gorilla: language acquisition in another pongid. *Brain and Language, 5 (1),* 72-97.

Ponce de León, M.S., Bienvenu, T., Marom, A., Engel, S., Tafforeau, P., et al. (2021). The primitive brain of early Homo. *Science, 372,* 165-171.

Read, D.W., Manrique, H.M., & Walker, M.J. (2022). On the working memory of humans and great apes: Strikingly similar or remarkably different? *Neuroscience & Biobehavioral Reviews, 134,* 104496.

Rizzi, L. (2016). Monkey morpho-syntax and merge-based systems. *Theoretical Linguistics 42 (1-2)*, 139-145.

Schlenker, P., Chemla, E., Arnold, K., & Zuberbühler, K. (2016). Pyow-hack revisited: Two analyses of Putty-nosed monkey alarm calls. *Lingua, 171,* 1-23.

Senghas, A., Kita, S., & Ozyurek, A. (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science, 305 (5691),* 1779-1782.

Stout, D., Hecht, E., Khreisheh, N., Bradley, B., & Chaminade, T. (2015). Cognitive demands of lower paleolithic toolmaking. *PLoS ONE 10 (4),* e0121804.

Wakita, M. (2020). Common marmosets (Callithrix jacchus) cannot recognize global configurations of sound patterns but can recognize adjacent relations of sounds. *Behavioural Processes, 176,* 104136.

Watumull, J., Hauser, M.D. Roberts, I.G., & Hornstein, N. (2013). On recursion. *Frontiers in Psychology, 4,* 1017.

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B, 375*, 20190062.

# PIECING TOGETHER THE BUILDING BLOCKS OF THE VOCAL LEARNING BAT BRAIN

NIENKE HOEKSEMA[*1,2], PETER HAGOORT[1], and SONJA VERNES[2]

[**]Corresponding Author: nienke.hoeksema@mpi.nl
[1] Neurobiology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[2]Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands & St. Andrews University, St. Andrews, Scotland

## 1. Background

Vocal learning is the ability to learn to produce novel vocalizations and is an important component of the human capacity to acquire speech and spoken language (Janik & Slater, 2000). Our close relatives, non-human primates, do not show evidence of learning to produce novel vocalizations. However, a small group of more evolutionarily distant animals, such as bats, display evidence for vocal learning. This raises the question of how and when this skill evolved and how it is encoded in our brains and genomes. *Phyllostomus discolor* is a highly social, vocal learning bat that can be studied in a laboratory setting (Vermes, 2017). Little is currently known about the brain areas and networks involved in vocal learning in *P. discolor*. Acquiring a better understanding of the neurobiology of this skill in bats will bring us closer to understanding how this skill evolved, and subsequently shed light on the evolution of human speech and spoken language.

## 2. Approach

In the current study, we used two complementary approaches, neuroimaging and neurogenetic mapping, to start to map the connectivity of the *P. discolor* brain and to highlight potential brain areas of interest for the study of vocal learning in *P. discolor*.

We performed high quality structural and multi-shell diffusion tensor imaging of post-mortem formalin-fixed *P. discolor* brains on an 11.7 T Bruker BioSpin magnetic resonance imaging (MRI) scanner. Previous research has shown increased and altered connectivity of brain areas involved in vocal communication in vocal learners compared to vocal non-learners (e.g., Kumar et al., 2016). As no research thus far has been done on investigating the macro-scale brain networks of *P. discolor*, we set out to exploratively map the brain-wide connectivity of *P. discolor*.

In parallel, we performed immunohistochemistry, immunofluorescence, and in situ hybridization on juvenile and adult *P. discolor* brains. We investigated a set of genes of which previous research has shown that they have a distinct pattern of expression in voluntary vocal motor (planning) regions in the brains of humans and vocal learning songbirds (Wang et al., 2015; Pfenning et al., 2014). We hypothesized that these genes might show the same pattern of expression in potential homologous bat vocal motor (planning) cortical regions, and could hence point us to potential brain regions of interest.

## 3. Preliminary Results and Future Perspectives

We successfully acquired high-quality DTI data of post-mortem *P. discolor* brains and are currently in the process of reconstructing the major white matter tracts in the *P. discolor* brain and performing probabilistic tractography from two cortical regions involved in social communication in *P. discolor*: the auditory and frontal cortex. Via neurogenetic mapping, we show that expression of our set of marker genes is not dispersed uniformly across the *P. discolor* cortex, but is enriched or reduced in specific layers and cortical areas. This suggests that these expression patterns can be used to mark out distinct brain areas in *P. discolor*.

Exploring our dataset further in the future, we hope to define key areas and circuits involved in vocal learning in the *P. discolor* bat brain and that this teaches us about how this skill evolved in bats and humans.

## References

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1–11. https://doi.org/10.1006/anbe.2000.1410

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., Howard, J. T., Wirthlin, M., Lovell, P. V., Ganapathy, G., Mouncastle, J., Moseley, M. A., Thompson, J. W., Soderblom, E. J., Iriki, A., Kato, M., Gilbert, M. T. P., Zhang, G., Bakken, T., … Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*(6215). https://doi.org/10.1126/science.1256846

Kumar, V., Croxson, P. L., & Simonyan, K. (2016). Structural Organization of the Laryngeal Motor Cortical Network and Its Implication for Evolution of Speech Production. *Journal of Neuroscience*, *36*(15), 4170–4181. https://doi.org/10.1523/JNEUROSCI.3914-15.2016

Vernes SC. What bats have to say about speech and language. Psychon Bull Rev. 2017 Feb;24(1):111-117. https://doi: 10.3758/s13423-016-1060-3

Wang, R., Chen, C. C., Hara, E., Rivas, M. V., Roulhac, P. L., Howard, J. T., Chakraborty, M., Audet, J. N., & Jarvis, E. D. (2015). Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners. *Journal of Comparative Neurology*, *523*(6), 892–906. https://doi.org/10.1002/cne.23719

# THE EFFECT OF IMPROVISATION AND LEARNING ON WORD ORDER PREFERENCES

ANNIE HOLTZ*[1], SIMON KIRBY[1], and JENNIFER CULBERTSON[1]

*Corresponding author: annie.holtz@ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, United Kingdom

**Background** In this study we hypothesise that certain typological patterns may have evolved as a result of competition between cognitive biases that tend to be active during different stages of language evolution. Research has shown that some cognitive biases influence linguistic behaviour under conditions where language structure is improvised (Culbertson, Schouwstra, & Kirby, 2020; Goldin-Meadow, So, Özyürek, & Mylander, 2008; Schouwstra & Swart, 2014), whereas other biases influence learning behaviour, meaning some linguistic structures are favoured or disfavoured during learning (Culbertson, Smolensky, & Legendre, 2012; Fedzechkina, Jaeger, & Newport, 2011; Hudson Kam & Newport, 2009). This research can be thought to represent two stages of language evolution, such that the biases that affect improvisation are active during language creation, and biases that influence learning affect language evolution only once there is a linguistic system for learners to acquire. However, recent research suggests that the influence of some biases may persist from improvisation into learning tasks (Motamedi, Wolters, Naegeli, Schouwstra, & Kirby, 2021). The possibility of these biases being active simultaneously means that competition between different biases could shape language typology.

We focus our investigation on a possible example of such competition, namely an exception to the typological tendency towards harmony, whereby languages tend to order dependents on the same side of the head (Greenberg, 1963; Hawkins, 1990). The exception concerns how adjectives (e.g. '*big house*') show a typological tendency for postnominal order (N-Adj = 879 vs Adj-N = 373, Dryer, 2013a) whereas genitives (e.g. '*The child's toy*') tend to be prenominal (N-Gen = 468 vs Gen-N = 685, Dryer, 2013b), despite a harmonic tendency among other nominal modifiers. We report results from an experiment where participants had to chose an order to expressing descriptive (adjective) and possessive (genitive) meanings in the absence of a language system. The second experiment explores if the ordering preferences identified in the first experiment continue to shape participants' linguistic behaviour in a learning task, where previous research has found that participants tend to have a preference for harmonic orders (Culbertson et al., 2012).

**Experiment 1** Participants in experiment 1 were randomly assigned to either an *adjective* condition (N=160) or a *genitive* condition (N=160). They saw images signifying meanings such as *striped book* (adjective condition) or *cyclops' hat* (genitive condition) accompanied by two gesture videos. In one of the videos the meaning of the image was conveyed with a prenominal gesture order (adjective/genitive-noun) and in the other with a postnominal gesture order (noun-adjective/genitive). Participants each saw a single such trial, and were instructed to choose the video they thought best conveyed the meaning of the image. Results from this experiment show that participants preferences generally align with those seen in typology, namely that they prefer postnominal orders for adjectives ($\beta$ = 0.51, SE. = 0.16, $z$ = 3.02, $p < 0.01$) and prenominal orders for genitives ($\beta$ = 0.56, SE = 0.16, $z$ = 3.43, $p < 0.001$).

**Experiment 2** In experiment 2 participants were randomly assigned to one of four conditions which differed in the word order used to convey adjective and genitive meanings. Each condition had a majority order shown 75% of the time, and a minority word order shown 25% of the time during training for each dependent type. In the natural condition the majority order for genitive trials was prenominal and for the adjective trials it was postnominal (i.e., following the preferences identified the typology, and replicated in experiment 1). The unnatural condition was the opposite of this. In the remaining two conditions the majority order was shared across both modifier types–either majority pre-nominal or post-nominal–and thus was harmonic. In the test phase, participants were shown an image to be conveyed along with two videos featuring pre- or post-nominal order and had to choose which they preferred, as in experiment 1. The results showed that participants generally learned the majority orders they were trained on ($\beta$ = 1.34, SE = 0.12, z = 11.24, p $<$ 0.001). Further, there was an overall tendency to regularise (i.e., use one order for a given dependent type more consistently than in the input) which was captured as reduction in conditional entropy ($\beta$ = -0.10, SE = 0.02, t = -8.74, p $<$ 0.001) and a tendency to harmonise (i.e., choose the same order across the two modifier types) which was measured as reduction in Shannon entropy ($\beta$ = -0.22, SE = 0.02, t = -5.15, p $<$ 0.001). Crucially, there was no evidence that the preferences identified in experiment 1 for specific modifier types continued to influence people's linguistic behaviour in this learning task ($\beta$ = -0.14, SE = 0.10, z = -1.40, p = 0.16). For example, participants did not learn more successfully, or regularise the majority order more, in the natural condition.

**Discussion** While participants showed a strong preference for postnominal adjectives and prenominal genitives when the task involved choosing an expression of descriptive and possessive meanings, this preference did not carry over to the learning task in experiment 2. The results suggest that this exception to harmony is not be caused by continuous competition between biases during learning. Future studies will examine if competition between these biases exist in language tasks which involve both improvisation and learning, such as extrapolation tasks.

## References

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: deriving biases in noun phrase order from statistical properties of the world. *Language*, *96*(3), 696–717.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329.

Dryer, M. S. (2013a). Order of adjective and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.*

Dryer, M. S. (2013b). Order of genitive and noun. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.*

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2011). Functional biases in language learning: Evidence from word order and case-marking interaction. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society.* Boston, MA: Cognitive Science Society.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163–9168.

Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language.* Cambridge, MA: MIT Press.

Hawkins, J. A. (1990). A parsing theory of word order universals. *Linguistic Inquiry*, *21*(2), 223–261.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, *59*(1), 30–66.

Motamedi, Y., Wolters, L., Naegeli, D., Schouwstra, M., & Kirby, S. (2021). Regularisation, systematicity and naturalness in a silent gesture learning task. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society.* Vienna, Austria: Cognitive Science Society.

Schouwstra, M., & Swart, H. de. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431–436.

# WORD ORDER CONVERGENCE IN THE HISTORY OF ENGLISH

MICHIO HOSAKA*[1]

*Corresponding Author: hosaka.michio@nihon-u.ac.jp
[1]The Department of English Language and Literature, Nihon University, Tokyo, Japan

## 1. Introduction

It is well known that early English had various types of word order, such as SXV, SVX, and XVS. However, in the course of the historical changes that have taken place in English, its word order has gradually come to be fixed as SVX. Such a word order change has been the target of historical studies for ages (van Kemenade 1987, Roberts 2007, inter alia). This paper focuses on the word order of the three-verb clusters (Modal+*have/be*+Participle) in OE and ME, and argues how it has converged in the history of English.

## 2. Methods and Data Analysis

This paper aims to clarify the word order change based on the statistical analysis of the data from the YCOE and PPCME2, and explore the mechanisms of its convergence from the perspective of the cultural evolution of language. As Table 1 shows, the word orders of 1-3-2 and 3-2-1 are the typical orders of Germanic languages and have decreased in number. On the contrary, the 1-2-3 order, which is the norm in PDE, has increased in the later period of OE. In addition, as Table 2 shows, it is obvious that the 1-2-3 order has been firmly established in ME.

| | Early OE | Late OE |
|---|---|---|
| 1-2-3 | 113 (55.4%) | 478 (86.8%) |
| 1-3-2 | 44 (21.6%) | 30 (5.4%) |
| 3-2-1 | 42 (20.6%) | 38 (6.9%) |
| 3-1-2 | 4 (2.0%) | 3 (0.5%) |
| 2-1-3 | 1 (0.5%) | 2 (0.4%) |
| 2-3-1 | 0 (0.0%) | 0 (0.0%) |
| Total | 204 | 551 |

Table 1 Three-verb clusters in OE

| | Early ME | Late ME |
|---|---|---|
| 1-2-3 | 365 (90.1%) | 1628 (98.4%) |
| 1-3-2 | 34 (8.4%) | 1 (0.1%) |
| 3-2-1 | 0 (0.0%) | 0 (0.0%) |
| 3-1-2 | 6 (1.5%) | 24 (1.5%) |
| 2-1-3 | 0 (0.0%) | 0 (0.0%) |
| 2-3-1 | 0 (0.0%) | 1 (0.1%) |
| Total | 405 | 1654 |

Table 2 Three-verb clusters in ME

Notes: 1-2-3 (Modal-*be/have*-Participle), 1-3-2 (Modal-Participle-*be/have*), 3-2-1 (Participle-*be/have*-Modal), 3-1-2 (Participle-Modal-*be/have*), 2-1-3 (*be/have*-Modal-Participle), 2-3-1 (*be/have*-Participle-Modal)

It has long been disputed philologically and theoretically as to why such word order changes have occurred in the history of English. The explanation can be found in the rise of functional projections (FPs) under the evolutionary assumption of dynamic structures being flexibly adaptive to communicative needs.

## 3. Discussion

Given that lexical projections are head-final and functional projections are head-initial, the structures of major three-verb clusters can be illustrated in the manner they have been shown in Figure 1.

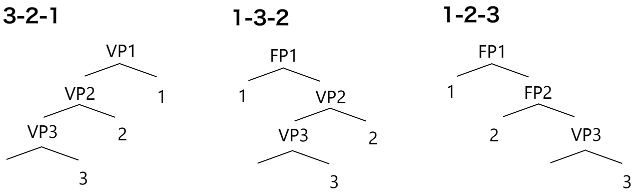


Figure 1 The derivation of major three-verb clusters in OE and ME

This shows that the word order of the three-verb clusters converged into 1-2-3 with the emergence of new functional projections. This can be regarded as the result of the rise of auxiliary verbs often explained in grammaticalization, which is assumed to be one of the processes of cultural evolution (Heine and Kuteva 2007, Smith 2011, inter alia). To be precise, ModalP, PassiveP and PerfectP emerged adaptively in order to compensate for the loss of verbal inflections in the course of the historical changes that have taken place in English. This convergence of word order is regarded as a good example of cultural evolution.

## References

Heine, B. and T. Kuteva (2007) *The Genesis of Grammar*. Oxford: Oxford University Press.

Kemenade, A. van (1987) *Syntactic Case and Morphological Case in the History of English*. Dordrecht: Foris.

Roberts, I. (2007) *Diachronic Syntax*. Oxford: Oxford University Press.

Smith, A.D.M. (2011) Grammaticalization and language evolution. In H. Narrog and B. Heine (eds.) *The Oxford Handbook of Grammaticalization*. 142-152. Oxford: Oxford University Press.

# TO WHAT EXTENT CAN THE HUMAN AGENT BIAS BE CAPTURED BY LINGUISTIC EXPERIENCE? EVIDENCE FROM MACHINE-LEARNING OF CORPORA

Eva Huber*[1,2], Paola Merlo[3], and Balthasar Bickel[1,2]

*Corresponding Author: eva.huber@uzh.ch
[1]Department of Comparative Language Science, University of Zurich.
[2]Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich.
[3]Département de Linguistique, University of Geneva.

Converging evidence from event apprehension (Wilson, Papafraou, Bunger, & Trueswell, 2011), language processing (Ferreira, 2003; Bornkessel-Schlesewsky & Schlesewsky, 2009; Kemmerer, 2012) and default grammar rules (Dryer, 2013; Napoli & Sutton-Spence, 2014) shows that humans have a bias towards the Agent participant in events (*Agent bias*). We contrast two hypotheses on its evolution: (i) The *biologically-driven hypothesis* ascribes the bias to an innate property that is already present in prelinguistic infants (Durrant et al., 2021) and that is potentially shared with other animals (Mascalzoni, Regolin, & Vallortigara, 2010); in this case, the effects on language stem directly from an innate principle. (ii) The *experience-driven hypothesis* ascribes the bias to our experience with the distributional properties in language use; in this case, the bias in processing is acquired, and the parallels between cognition and language reflect parallel evolution. To probe these hypotheses we focus on the specific effect that Agent Bias has on the processing of role-ambiguous noun phrases. Here the effect reveals itself in a transient prediction of an agent role which triggers an electrophysiological deflection, typically an N400 effect, when the prediction fails later in the sentence (Bornkessel-Schlesewsky & Schlesewsky, 2014). The experience-driven hypothesis suggests that this effect can be simulated with computational models whose knowledge of language is derived from distributional patterns in the linguistic input and general architectural assumptions but without any explicit agent bias built into them. Previous work on English (Levy, 2008; Frank, Otten, Galli, & Vigliocco, 2015; Szewczyk & Federmeier, 2022) has shown that N400 effects correlate with *surprisal*, the negative log probability of a word given the previous input extracted from such models. We use previously conducted EEG experiments which contain an initial ambiguous NP to compare the extent to which suprisal values in these models can capture N400 effects. Under the experience-driven hypothesis, we expect higher surprisal values in the conditions that also trigger an

N400 effect. Under the biologically-driven hypothesis, we expect higher surprisal values not to correlate with the conditions that raise an N400 effect.

We selected experimental data from three languages that are maximally distinct in how they code Agents: Hindi (Bickel, Witzlack-Makarevich, Choudhary, Schlesewsky, & Bornkessel-Schlesewsky, 2015), which flags Agents by an "ergative" case maker only under specific conditions (perfective transitive clauses); Basque (Isasi-Isasmendi Andaluze et al., 2022), which flags all Agents as ergative; and German (Haupt, Schlesewsky, Roehm, Friederici, & Bornkessel-Schlesewsky, 2008), which does not flag Agents by case but treats them as general subjects. All experiments showed that a sentence-initial noun phrase disambiguated to a Patient reading elicits an N400 response, although experimental conditions varied (Hindi: role-ambigous vs unambigous Patients × aspect; Basque: ambiguous Agents vs Patients × unambigous Agents vs Patients; German: ambigous Agents vs Patients × dative vs accusative-assigning verbs).

We extracted the surprisal values of two types of language model architectures: a Long Short-Term (LSTM) model (Hochreiter & Schmidhuber, 1997), a variant of a recurrent neural network, and a Transformer model (Vaswani et al., 2017). LSTMs process units incrementally by recurrence, and hence contain a memory bottleneck. Transformers, on the contrary, have 'attention layers' which allow for direct access to parts of the previous input (Vaswani et al., 2017). Hierarchical Bayesian Models are fitted to estimate surprisal values given the same conditions as in the experiments. We then qualitatively evaluate whether the surprisal values show the same trends as the N400, in line with the experience-driven hypothesis.

The results (see figure in supplementary material) for the Hindi LSTM show a slight trend in line with the N400 findings, but estimates overlap between conditions (89% CIs imperfective role-ambiguous: [13.29,14.55] vs role-marked: [12.9,14.12], perfective amb.: [11.57,12.74] vs marked: [11.24,12.31]); the Transformer model replicates the N400 findings slightly better in the imperfective aspect (role-marked [16.53,17.88] vs role-unmarked [15.74,17.02]), unlike in the experiments, where it held across aspects. Both Basque models replicates the N400 findings (LSTM A: [12.55,13.61 vs P: [13.63,14.47]; Transformer A: [16.18,16.26] vs P: [16.24,16.32]) but, unlike in the experiments, they generalize the pattern to unambigous cases (LSTM A: [12.71,13.78] vs P: [13.44,14.29]; Transformer: [16.18,16.26] vs P: [16.23,16.31]), where no N400 was observed. The German models replicate the N400 findings but the effect is strong and reliable only in the Transformer model (for acc-verbs: A: [3.31, 4.1] vs P: [5.73,6.48]; LSTM: A: [8.96,9.76] vs P: [9.67,10.29]).

We conclude that there is at best marginal evidence for the experience-driven hypothesis. The surprisal values replicate N400 effects only in German Transformer models. In all other cases, surprisal values are either more general (Basque) or less general (Hindi) than N400 findings; in addition most effects have overlapping credibility intervals.

## Acknowledgments

This work was funded by the NCCR Evolving Language supported by Swiss National Science Foundation Agreement (#51NF40_180888).

## References

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLOS ONE*, *10*, e0132819.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Minimality as vacuous distinctness: Evidence from cross-linguistic sentence comprehension. *Lingua*, *119*(10), 1541–1559.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2014). Competition in argument interpretation: Evidence from the neurobiology of language. *Competing motivations in grammar and usage*, *107*, 126.

Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Durrant, S., Jessop, A., Chang, F., Bidgood, A., Peter, M. S., Pine, J. M., & Rowland, C. F. (2021). Does the understanding of complex dynamic events at 10 months predict vocabulary development? *Language and Cognition*, *13*(1), 66–98.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, *47*(2), 164–203.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and language*, *140*, 1–11.

Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., & Bornkessel-Schlesewsky, I. (2008). The status of subject–object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, *59*(1), 54–96.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780.

Isasi-Isasmendi Andaluze, A., Sauppe, S., Andrews, C., Laka, I., Meyer, M., & Bickel, B. (2022). Semantic role parsing in intransitives: EEG insights from basque. *34th Annual CUNY Conference on Human Sentence Processing*.

Kemmerer, D. (2012). The cross-linguistic prevalence of sov and svo word orders reflects the sequential and hierarchical representation of action in broca's area. *Language and Linguistics Compass*, *6*, 50-66.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Mascalzoni, E., Regolin, L., & Vallortigara, G. (2010). Innate sensitivity for self-propelled causal agency in newly hatched chicks. *Proceedings of the National Academy of Sciences*, *107*(9), 4483–4485.

Napoli, D. J., & Sutton-Spence, R. (2014). Order of the major constituents in sign languages: implications for all language. *Frontiers in psychology*, *5*.

Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of Memory and Language*, *123*, 104311.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wilson, F., Papafraou, A., Bunger, A., & Trueswell, J. (2011). Rapid extraction of event participants in caused motion events. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33).

# PHYLOGENETIC RECONSTRUCTION FOR JAPONIC: NEW DATA TYPES AND METHODOLOGICAL INSIGHTS

JOHN L. A. HUISMAN*[1], BONNIE MCLEAN[1], and CHIEH-HSI WU[2]

*Corresponding Author: john.huisman@lingfil.uu.se
[1]Department of Linguistics and Philology, Uppsala University, Uppsala, Sweden
[2]Mathematical Sciences, University of Southampton, Southampton, UK

## 1. Background

While synchronic variation in the Japonic languages is well understood from the meticulous work of Japanese dialectology, the relationships between the various linguistic groups from a historical perspective have received considerably less attention. Existing studies tend to overrepresent either of the two main branches of the family—i.e. Japanese (Lee & Hasegawa, 2011) and Ryukyuan (Pellard, 2009, 2015)—leaving some questions about the full family history unanswered. Bayesian phylogenetic analyses have the potential to recover these deeper-level branching events, but the complexity of such analyses requires large amounts of data, which have been argued to only be "really available from cognacy in the lexicon" (Greenhill et al., 2020). This can pose a challenge in contexts of recent diversification, areal features, and regular contact where lexical diversity is limited—e.g., the situation found on the Japanese mainland (Hattori, 1973).

## 2. The current study

To address these issues, we analysed a sample of languages that more effectively captures linguistic diversity of both the Japanese mainland and the Ryukyuan islands, and experiment with analyses that—unlike the prevalent approach in computational historical linguistics—include features other than cognate sets as phylogenetic characters. Specifically, we use phonotactic traits as phylogenetic characters using the method described in Macklin-Cordes et al. (2021), and assess their informativeness for language families with a relatively limited time-depth.

### *2.1. Data and methods*

We started by collating basic vocabulary data for 65 Japanese and 28 Ryukyuan varieties using a 250-item list compiled from other commonly used concept lists (Greenhill et al., 2008; Swadesh, 1952; Tadmor, 2009) and coded the lexemes for cognacy. Subsequently, we used the vocabulary data to extract phonotactic traits using the approach set out by Macklin-Cordes and colleagues (2021), i.e., the presence/absence of sequences of sound segments, the relative transition frequencies between sound segments, and the relative transition frequencies between sound classes. We inferred phylogenetic trees using Bayesian approaches (Bouckaert et al., 2019) with different tree priors and models of evolution in three analyses (lexical traits only, phonotactic traits only, and lexical and phonotactic traits combined), and compare our findings to previously suggested phylogenies using different methods.

### *2.2. Results*

As expected, we found that reduced levels of lexical diversity resulted in low clade support, particularly among the Japanese mainland varieties, in the analysis based on lexical traits alone. In the analysis on phonotactic traits alone, we found several (albeit minor) conflicts with generally accepted phylogenies for the Ryukyuan branch—which, we discuss, likely result from parallel changes in the phoneme inventory that reverberate in phonotactics. However, when combining the lexical and phonotactic traits into a single holistic analysis, the model was able to both capture language relationships accurately, while also considerably decreasing the overall uncertainty in the model. We provide an overview of the strengths of the method used here (e.g., "free" algorithmically inferred data), as well as the challenges that remain with potential solutions, e.g., testing against simulated data (Wichmann & Rama 2021). We also discuss how our results shed light on ongoing questions in Japanese historical linguistics (e.g., the overall tree topology; the timing of initial diversification), and previously raised concerns about how the spatial structure of the Japonic language family complicates phylogenetic inference and its interpretation (e.g., Murawaki, 2015).

## 3. Summary

We present a case study of the use of phonotactic traits in computational historical linguistics (following Macklin-Cordes et al. 2021). Our findings suggest that non-lexical data can be a valuable addition to analyses that aim to untangle fine-grained phylogenetic structures in contexts of recent diversification and language contact. The reconstructed dated phylogeny of the Japonic language family provides new insights into the linguistic history of Japan, which can e.g., further our understanding of geographic factors in language diversification (cf. Huisman et al., 2019), or facilitate assessing the validity of predictive mathematical models of language change (cf. Takahashi & Ihara, 2020).

## References

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., ... & Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology, 15*(4), e1006650.

Greenhill, S. J., Blust, R., & Gray, R. D. (2008). The Austronesian basic vocabulary database: From bioinformatics to lexomics. *Evolutionary Bioinformatics*, *4*, EBO–S893.

Greenhill, S. J., Heggarty, P., & Gray, R. D. (2020). Bayesian phylolinguistics. *The Handbook of Historical Linguistics*, *2*, 226–253.

Hattori, S. (1973). Japanese dialects. In H. Hoenigswald (Ed.), *Diachronic, areal and typological linguistics* (pp. 368–400). De Gruyter Mouton.

Huisman, J. L. A., Majid, A., & Van Hout, R. (2019). The geographical configuration of a language area influences linguistic diversity. *PLOS ONE*, *14*(6), e0217363.

Lee, S., & Hasegawa, T. (2011). Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1725), 3662–3669.

Macklin-Cordes, J. L., Bowern, C., & Round, E. R. (2021). Phylogenetic signal in phonotactics. *Diachronica*.

Murawaki, Y. (2015). Spatial structure of evolutionary models of dialects in contact. *Plos One*, *10*(7), e0134335.

Pellard, T. (2009). *Ōgami: Éléments de description d'un parler du Sud des Ryūkyū*. Ecole des Hautes Etudes en Sciences Sociales.

Pellard, T. (2015). The linguistic archeology of the Ryukyu Islands. In P. Heinrich, S. Miyara, & M. Shimoji (Eds.), *Handbook of the Ryukyuan Languages. History, Structure, and Use* (pp. 13–38). De Gruyter Mouton.

Swadesh, M. (1952). Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society*, *96*(4), 452–463.

Tadmor, U. (2009). Loanwords in the world's languages: Findings and results. *Loanwords in the World's Languages: A Comparative Handbook*, *55*, 75.

Takahashi, T., & Ihara, Y. (2020). Quantifying the spatial pattern of dialect words spreading from a central population. *Journal of The Royal Society Interface*, *17*(168), 20200335.

Wichmann, S., & Rama, T. (2021). Testing methods of linguistic homeland detection using synthetic data. *Philosophical Transactions of the Royal Society B: Biological Sciences, 376*(1824), 20200202.

# NON-LINEARITY IN HUMAN LANGUAGE EVOLUTION: A GENERAL ECOLOGY PERSPECTIVE GROUNDED IN PLEISTOCENE ARCHAEOLOGY

SHUMON T. HUSSAIN[1] and ALEXANDRA R. KRATSCHMER*[2]

*Corresponding Author: romak@cc.au.dk
[1]Department of Archaeology and Heritage Studies, Aarhus University, Aarhus, Denmark
[2]Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, Aarhus, Denmark

This paper reconsiders the emergence of language in the hominin lineage from the perspective of general ecology, multispecies theory, phenomenology, and semiotics. We propose that changing communicative environments and their shifting foci of hominin-animal and hominin-hominin interaction are fundamental to understanding the evolution of language in the hominin lineage. After reviewing the available contextual evidence on the nature and transformation of communicative ecologies in the deep-past, we pledge for interpretive conservatism and argue that an "extended short chronology" for the evolution of human language is currently best supported by the archaeological data (in sync with a prolonged period of intensified "indexical" communication). We further propose that this evolutionary trajectory is best understood as a form of non-linear "co-evolution" and hence interlaced with key developments in multiple hominin behavioral arenas and material registers, such as animal relationships, technical behavior, artistic expression and the modalities and scale of social life. Systematically mapping and exploring this broader "general ecology" of communicative needs, concerns, functions, and horizons minimally requires the integration of archaeology, ethology, linguistics, and phenomenology.

## 1. Introduction: Shifting the Theoretical Focus

Asking for the origins and evolution of human language is a long-standing conundrum. The available evidence is necessarily indirect – language does not "fossilize" – and assumptions about what precisely constitutes evidence for language in the first place – and more difficult even, language evolution – differ vastly across disciplinary and theoretical spectra. Accounts of the evolution of human language range from cognitive approaches, localizing language as an extension or foundation of generalized capacities for symbolic thought (e.g., Noble & Davidson, 1991; Donald 1991; Henshilwood & d'Errico, 2011;

Tattersall, 2019), over cultural semiotic, mostly Piercean, ones (Kissel & Fuentes, 2017; Barham & Everett, 2021) to ethologically-founded communication studies arguing for deep-historical roots of complex communication systems in diverse animal taxa, especially birds, and the co-evolution of "speech", tool-use, social behavior, and potentially aesthetic display, in a number of unrelated species (Menninghaus, 2011; Stout, 2018). Yet, the main sources of evidence for human evolutionary studies of language are material culture assemblages and specific types of artefacts uncovered by archaeologists and millennial-scale changes in hominin bio-cognitive makeup documented by paleoanthropologists, cognitive scientists, and primatologists (Dediu & Levinson, 2018). The Darwinian optic that goes together with these disciplinary configurations has vindicated the idea that the emergence of language – because it "makes us human" – is a watershed event or turning point in hominin evolution, dividing the past into a "before" and "after". The result has been a tendency to promote unilineal models of language evolution rooted in persistent notions of progress, leading from the first representatives of the *Homo* lineage to the only surviving hominin taxon, *H. sapiens*. We suggest that the problematic vision of a set language threshold that isolates a so-called "archaic" from a "modern" stage in human prehistory, where full-blown humanity is achieved, is linked to an overemphasis, and at times fetishization, of intrinsic "capacities" at the expense of *context* and *communicative ecologies* in hominin evolution.

Definitions and conceptual baselines: we here retain the conceptual distinction between "language" as a fully-fledged symbolic sign system in the Saussurian sense, where each sign obtains its meaning by virtue of its relation to other signs, and "communication" as "[a] process involving signalling between a sender and receiver, resulting in a perceptual response in the receiver, which extracts information from the signal, potentially influencing the receiver's behavior" (Stevens, 2013: 73). Communication in this view is a more basal and far-flung phenomenon than language and as such deeply involved in most life-processes and inter-organismic relations (Barbieri, 2008). Language can then be said to respond to specific communicative needs and serves particular functions, which may in turn also change the total architecture of broader *communicative ecologies* – i.e., the contexts and relations that frame communicative processes at large. This perspective shifts the attention to the social, cognitive, and environmental embeddedness of language as a specific form of communication rather than reducing it to a mere question of "language ability" (Lieberman, 2006).

## 2. The Big-Picture Perspective: Changing Niches and "Ecologies of Communication" over the last 2 Million Years

This paper reconsiders language evolution among hominins in the framework of general ecology, multispecies theory, phenomenology and biosemiotics (meaning-making in living organisms). We contend that previous approaches have paradoxically endorsed a human-centered point of departure with often quasi-orthogenetic implications. Our advocated approach flips this perspective on its head. We draw attention to the fact that early hominins were an extraordinarily rare sights in Plio-Pleistocene landscapes, and in turn primarily had to grapple with nonhuman-dominated life-worlds (Rodríguez et al., 2015). The communicative needs and concerns in these contexts are fundamentally *multispecies*: early hominin foraging niches relied on the coordination with other animals, especially predators and key scavengers, and biosemiotic relationships at the hominin-animal interface were thus of key existential concern. Communication systems therefore likely evolved in tandem with such broader interspecies niche conditions as well as the changing demographic and social contexts hominins found themselves in. Drawing on ethological data demonstrating structured relationships between different niche-types and communicative environments, we explore the consequences of broad "human niche" trajectories – with spatiotemporally divergent patterns of evolving carnivory and generalist foraging (Ben-Dor et al., 2020) – and highlight the importance of the Middle Pleistocene (ca. 600-130,000 years ago) associated with flourishing, wide-spread archaeological evidence for novel forms of social organization rooted in "soft" communicative technologies such as hearth-centered life, meat sharing and big-game hunting (Kuhn & Stiner, 2019).

The Middle Pleistocene not only documents important changes in the hominin body plan including marked expansions in cranial capacity, but the archaeological record also points to significant reconfigurations of hominin social and technical ecologies. At Qesem cave in modern-day Israel, hominins were roasting and dining on tortoise shells about 400,000 years ago and differential ungulate bone representation in the same layers indicates selective movement of meat-rich body parts to "homebase"-like places equipped with hearths (Stiner et al., 2011). Such places not only promote and intensify social interaction among peers, but they also generate novel contexts for the sharing of experience and what has been referred to as "gossiping" (Dunbar, 2017), and thus ultimately precipitate the evolutionary dynamics between hominin place-making and storytelling resulting in the diagnostic “storied” landscapes documented among contemporary forager groups. Evidence for the emergence of hafted spear tips used for hunting larger animals, although sporadic, also reaches back ca. 500,000 years (Wilkins et al., 2012). Generally speaking, the Middle Pleistocene sees the proliferation, diversification and formalization of so-called “prepared core technologies” based on the principle of anticipated morphotechnical control over stone knapping products as well as different forms of technical “predetermination”, often with a notable “division of technical labor” among varying coexisting technologies.

This goes hand in hand with the development of (non-linear) production trajectories comprising “intermediate” products which in principle allow for the separation of blank (unmodified artefacts) and stone tool production chains. This phenomenon is for example signaled by erupting evidence for tool hafting at the end of the Middle Pleistocene, and paves the ground for task differentiation in both tools and tool users and thus the promulgation of “communities of practice” with new requirements for social coordination beyond the spatiotemporal horizon of the “here and now“. We propose that such developments have re-directed the communicative foci of hominins from the wider more-than-human world to their hominin peers. The development of taxing stone technologies such as “Levallois”, which depend on faithful horizontal and vertical transmission of technical knowledge and its cross-generational stabilization in the timeframe between 400 and 300,000 years ago has, consistent with this view, recently been interpreted as evidence for a new quality of hominin intergroup interaction and sociotechnical information exchange (MacDonald et al., 2021).

The Middle-to-Late Pleistocene transition (ca. 300-100,000 years ago) corresponds to the crystallization of new hominin behaviors linked to the collaborative stalking and intercepting of isolated large animals such as rhinos and bears or entire groups or even herds of animals such as horse or reindeer. The implicated hunting tactics have been argued to require planning and cooperation, and hence reliable intragroup communication, even though collective hunting is generally consistent with “language-free” explanations, as for example shown by lions, orcas and many other animals exerting cooperative foraging. Yet, it is notable that the Middle-to-Late Pleistocene interface is also linked to a mosaic of varying hominin taxa populating different landscapes, and possibly coexisting in some, providing a new context of “multispecies” interaction and communication (Wood & Boyle, 2016). That these different hominin forms at least occasionally encountered and interacted with each other is demonstrated by aDNA research

(Liu et al., 2021). This “pluralism” of hominin contexts and their overlap or intersection as well as the coercing climatic envelope of the developing Middle Pleistocene bringing about high-amplitude 100,000-year warm-cold cycles with often rapid transitions speaks to the dynamic and unsteady character of the associated communication ecologies, so that hominin extinctions and dislocations were probably not the exception (Hussain & Riede, 2020). The diverse physical anthropological record of Eastern Eurasia and Africa not only suggests amplified gene-culture coevolution but simultaneously points to novel and increasingly existential requirements of navigating landscapes frequented by other hominins, and to communicatively attend to them.

As hominins become more frequent on the landscape – starting perhaps as early as 100,000 years ago in some areas – and as they begin to increasingly rely on each other across vast geographic distances, communicative ecologies presumably change once more and communication transcends its sole information-conveyor service, developing into a medium of cultural signposting and identity-formation within interconnected human-concentrated worlds (Roberts, 2013). Anthropological and philological observations indicate that such communicative ecologies are prone to promote multi-linguality, and we may surmise that such conditions were in place at least since the Late Pleistocene. The Late Pleistocene (ca. 125-11,000 years ago), non-coincidentally, is associated with the emergence of forms of visual culture indicating a consequential role of material objects in grounding social memory and relaying storied worlds (Porr, 2010). The development of broadly delimited cultural geographies rooted in distinct artefact designs (lithic tools, personal ornaments, ochre and ostrich shell engravings) has traditionally been linked to this timeframe as well (Vanhaeren & d’Errico, 2006; Brooks et al., 2006; Ruebens, 2013; Tylén et al., 2020; Hussain & Will, 2021). At the end of the Pleistocene, stone artefact forms begin to enact group awareness and possibly embody what Wiessner (1983) has called “emblematic style”. Bow-and-arrow technology which, based on current knowledge, makes its first appearance in the archaeological record around 70,000 years ago has been argued to depend on complex technical knowledge and “delayed” problem-solving (Lombard & Haidle, 2012) and thus similarly supports the idea that both the contents and the contexts of hominin communication were subjected to transformative change in the Late Pleistocene.

We suggest it is no coincidence that such social learning-dependent technologies flourish in the context of “generalist” hominin niches which depart from earlier carnivory-reliant hominin foraging strategies (Roberts & Steward, 2018; Ben-Dor et al., 2021). Such “niche broadening” at the end of the Pleistocene has important consequences for ecologies of communication since foragers rely on a deeper and broader vocabulary of places and resources, which promotes the development of differentiated zoo- and pythotaxonomies and has implications for the communicative needs of task allocation.

## 3. Conclusion: Non-Linearity, Interpretative Conservativism and an “Extended Short Chronology” of Language in the Hominin Lineage

Searching for the origins of human language in the archaeological and paleoanthropological records may be a futile exercise since it simply remains unclear what constitutes unambiguous evidence for language *vis-à-vis* merely indexical communication. Perhaps more importantly, the lines of evidence traditionally linked to language evolution are greatly staggered in time, rather than being correlated or chronologically “packed”, and thus suggest a gradual language *continuum* rooted in changing ecological conditions, rather than a set of mysterious hominin predispositions. The indirect archaeological evidence available is additionally patchy and thus generally supports a non-linear evolution of hominin communicative environments instead of a single context of language “origin”. Given these conditions, we pledge for interpretive conservatism and conclude that an “extended short chronology” for the evolution of language in the hominin lineage is currently best supported by the archaeological evidence (in sync with a prolonged period of “indexical” communication). We propose that this evolution is best understood as a form of “co-evolution”, and thus tied to key developments in multiple behavioral arenas and material registers such as hominin-animal interaction, technical behavior, artistic expression, and social life. Systematically mapping and exploring this broader “general ecology” of communicative needs, concerns and functions requires at least the integration of archaeology, ethology, linguistics, semiotics, and phenomenology. This coupled perspective makes room for variously emerging and collapsing ecologies of communication without losing sight for the general lines of development in hominin evolution. The result is a more realistic, yet intricate picture of human language evolution giving rise to an original compromise between radical “short” and “long” chronology renderings of language evolution.

## References

Barbieri, M. (2008). Biosemiotics: a new understanding of life. *Naturwissenschaften*, *95 (7),* 577-599.

Barham, L., & Everett, D. (2021). Semiotics and the origin of language in the Lower Palaeolithic. *Journal of Archaeological Method and Theory*, *28 (2),* 535-579.

Ben-Dor, M., Sirtoli, R., & Barkai, R. (2020). The evolution of the human trophic level during the Pleistocene. *Yearbook of Physical Anthropology*, *175,* 27-56.

Brooks, A. S., Nevell, L., Yellen, J. E., & Hartman, G. (2006). Projectile Technologies of the African MSA. In E. Hovers & S. L. Kuhn (Eds.), *Transitions Before the Transition* (pp. 233-155). Boston: Springer.

Dediu, D., & Levinson, S.C. (2018). Neanderthal language revisited: not only us. *Current Opinion in Behavioral Sciences*, *21,* 49-55.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard UP. Print.

Dunbar, R. I. M. (2017). Group size, vocal grooming and the origins of language. *Psychonomic Bulletin* & *Review*, *24,* 209-212.

Henshilwood, C. S., & d'Errico, F. (Eds.) (2011). *Homo symbolicus: the dawn of language, imagination and spirituality*. Amsterdam: John Benjamins Publishing.

Hussain, S. T., & Riede, F. (2020). Paleoenvironmental humanities: Challenges and prospects of writing deep environmental histories. *Wiley Interdisciplinary Reviews: Climate Change, 11 (5),* e667.

Hussain, S. T., & Will, M. (2021). Materiality, Agency and Evolution of Lithic Technology: An Integrated Perspective for Palaeolithic Archaeology. *Journal of Archaeological Method and Theory*, *28 (2),* 617-670.

Kissel, M., & Fuentes, A. (2017). Semiosis in the Pleistocene. *Cambridge Archaeological Journal*, *27 (3),* 397-412.

Kuhn, S. L., & Stiner, M. C. (2019). Hearth and Home in the Middle Pleistocene. *Journal of Anthropological Research*, *75 (3),* 305-327.

Lieberman, P. (2006). *Toward an Evolutionary Biology of Language*. Yale: Belknap.

Liu, Y., Mao, X., Krause, J., & Fu, Q. (2021). Insights into human history from the first decade of ancient human genomics. *Science, 373*, 1479-1484.

Lombard, M., & Haidle, M. N. (2012). Thinking a Bow-and-Arrow Set: Cognitive Implications of Middle Stone Age Bow and Stone-tipped Arrow Technology. *Cambridge Archaeological Journal, 22 (2),* 237-264.

MacDonald, K., Scherjon, F., van Veen, E., Vaesen, K., & Roebroeks, W. (2021). Middle Pleistocene fire use: The first signal of widespread cultural diffusion in human evolution. *Proceedings of the National Academy of Sciences of the United States of America, 118 (31),* e2101108118.

Menninghaus, W. (2011). *Wozu Kunst? Aesthetik nach Darwin*. Berlin: Suhrkamp.

Noble, W., & Davidson, I. (1991). The evolutionary emergence of modern human behaviour: Language and its archaeology. *Man*, *26 (2),* 223-253.

Porr, M. (2010). Palaeolithic Art as Cultural Memory: A Case Study of the Aurignacian Art of Southwest Germany. Cambridge *Archaeological Journal, 20 (1),* 87-108.

Roberts, G. (2013). Perspectives on Language as a Source of Social Markers. *Language and Linguistics Compass*, *7 (12),* 619-632.

Roberts, P., & Stewart, B.A. (2018). Defining the 'generalist specialist' niche for Pleistocene Homo sapiens. *Nature Human Behaviour*, *2 (8),* 542-550.

Rodríguez, J., Mateos, A., Martín-González, J. A., & Rodríguez-Gómez, G. (2015). How rare was human presence in Europe during the Early Pleistocene? *Quaternary International, 389*, 119-130.

Ruebens, K. (2013). Regional behaviour among late Neanderthal groups in Western Europe: a comparative assessment of late Middle Palaeolithic bifacial tool variability. *Journal of Human Evolution*, 65 *(4),* 341-362.

Stevens, Martin. 2013. *Sensory Ecology, Behavior and Evolution*. Oxford: Oxford University Press.

Stiner, M. C., Gopher, A., & Barkai, R. (2011). Hearth-side socioeconomics, hunting and paleoecology during the late Lower Paleolithic at Qesem Cave. Israel. *Journal of Human Evolution, 60 (2),* 213-233.

Stout, D. (2018). Archaeology and the evolutionary neuroscience of language: The technological pedagogy hypothesis. *Interaction Studies*, *19 (1/2),* 256-271.

Tattersall, I. (2019). The Minimalist Program and the origin of language: a view from paleoanthropology. *Frontiers in Psychology*, *10*. 677.

Tylén, K., Fusaroli, R., Rojo, S., Heimann, K., Fay, N., Johannsen, N. N., Riede, F., & Lombard, M. (2020). The evolution of early symbolic behavior in Homo sapiens. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(9), 4578-4584. https://doi.org/10.1073/pnas.1910880117

Vanhaeren, M., & d'Errico, F. (2006). Aurignacian ethno-linguistic geography of Europe revealed by personal ornaments. *Journal of Archaeological Science, 33 (8),* 1105-1128.

Wiessner, P. (1983). Style and Social Information in Kalahari San Projectile Points. *American Antiquity*, *48*, 253-276.

Wilkins, J., Schoville, B. J., Brown, K. S., & Chazan, M. (2012). Evidence for Early Hafted Hunting Technology. *Science, 338*, 942-946.

Wood, B., & Boyle, E. K. (2016). Hominin taxic diversity: Fact or fantasy? *American Journal of Physical Anthropology, 159,* 37-78.

# GENERATION OF STRUCTURED SPATIAL PATTERNS THROUGH ADVERSARIAL IMITATION LEARNING

Hiroyuki Iizuka*[1,2], Wataru Noguchi[1], Taiki Sasaki[3]and Masahito Yamamoto[1,2]

*Corresponding Author: iizuka@ist.hokudai.ac.jp
[1]Faculty of Information Science and Technology, Hokkaido University, Sapporo, Japan
[2]Center for Human Nature, Artificial Intelligence, and Neuroscience,
Hokkaido University, Sapporo, Japan
[3]Graduate School of Information Science and Technology,
Hokkaido University, Sapporo Japan

This study shows that systematic structures that appear in language and birdsong can be generated by adversarial imitation learning, in which one wants to imitate the opponent's patterns without the opponent imitating their patterns. In prior adversarial imitation studies, the generated patterns became chaotic but not structured. Therefore, we extended a previous model using the concept of generative adversarial networks (GANs) for deep learning. The agents were modeled using generator and discriminator networks. Our results show that mutual adversarial imitation learning can lead to higher fractal dimensions of the generated patterns and cause the structurization of patterns.

## 1. Introduction

Aside from being a complex time-series pattern, a language also features a systematic structure or grammar (Chomsky, 1965). Such a structure can be observed in the time-series of the bird songs (Honda & Okanoya, 1999), the close calls of banded mongoose (Jansen et al., 2012) and the humpback whale songs (Cholewiak et al., 2013) as well as in the behavior patterns of macaque monkeys (Hihara et al., 2003) and degus (Tokimoto & Okanoya, 2004) and even in spatial patterns (Kondo & Miura, 2010).

To determine how systematic structures emerged, Kirby et al. conducted experiments from a linguistic perspective (Kirby et al., 2008, Cornish et al., 2013, Winters et al, 2015). These experiments showed that linguistic structures emerge in the trade-off between learnability and expressivity through cultural transmission, in which symbols are remembered and then transmitted to others repeatedly (remembering here is equivalent to what we call imitation). However, these experiments required the assignment of distinct random sequences for each expression at the initial stage in order to produce structured sequences. If the initial sequences are equivalent among all expressions, the sequence is simply transmitted without any structuring, and the system has no expressivity. This means that other principles are needed to explain how the complexity required to produce structural patterns emerges.

Accordingly, we hypothesize that patterns generated by organisms become complex and structured through adversarial situations, in which one wants to imitate or understand an opponent’s signal patterns without the opponent imitating or understanding their own signal patterns. For example, the ability to imitate the profitable behavioral patterns of others leads to one's own benefits. In contrast, this may remove these benefits from those who are being imitated. In other words, the strategy of imitating others while avoiding being imitated by others may be evolutionarily dominant. Furthermore, individuals may want only those in their in-group to understand profitable signals, which requires secrecy. We believe that the complexity generated by interspecific adversarial communication is adopted for intraspecific cooperative communication, which in turn leads structured communication to evolve.

Based on this concept, we propose a simulation model that complexifies time-series patterns through adversarial imitation learning. Although the original model only produced chaotic behavior in time-series (Yamazaki et al., 2020), we extended it by introducing the recognition of patterns by individuals, thus showing that recognition leads to simplification and pattern structuring, rather than mere complexity.

## 2. Simulation model

In the field of machine learning, generative adversarial networks (GANs) can imitate real data (e.g., images) to generate realistic artificial data (Goodfellow et al., 2014). GANs produce high-quality imitation data using a generator and a discriminator that features adversarial learning. We used GANs to model a situation in which two individuals, Agents 1 and 2, are engaged in adversarial imitation learning. Unlike the study upon which our work is founded, we introduced the recognition of a discriminator to evaluate the imitation’s success (Suzuki & Kaneko, 1994, Yamazaki et al., 2020).

An overview of the proposed model is presented in Fig. 1. Each individual consists of a generator (G1/G2) and a discriminator (D1/D2). The discriminator determines whether a pattern (G1(Z) or G2(Z)) is produced by its own generator or the opponent’s generator. The generator learns to generate a pattern such that its own discriminator can easily recognize the pattern as being generated by itself. This learning process works toward simplifying the pattern to facilitate recognition. Furthermore, the generator learns to trick the opponent discriminator by recognizing patterns produced by the opponent generator. Consequently, the discriminator and generator learn to avoid being tricked by the opponent’s generator, thus increasing pattern complexity. This learning process occurs under mutual conditions.

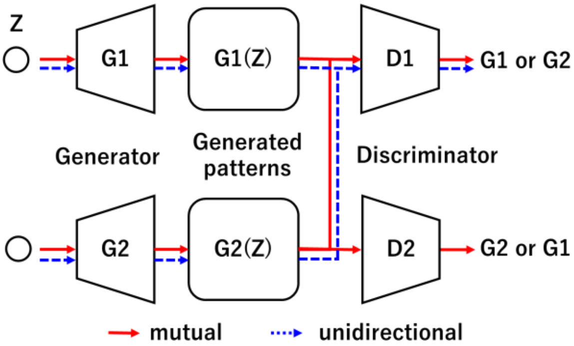


Figure 1. Overview of our model. G1/G2 and D1/D2 represent the generators and discriminators for Agent 1/2, respectively. Red and blue arrows indicate mutual and unidirectional adversarial imitation learning, respectively. The generators and discriminators are realized by deep neural networks with convolutions and de-convolutions. The generators receive random values as seeds to generate patterns, as in the original GAN. G1/2(Z) are the patterns created by Generators G1/2.

Both the generator and the discriminator are implemented with feed-forward neural networks, with initial parameters given randomly. Those parameters are trained by gradient descent. The loss function of Agent 1 can be expressed using hinge loss as follows:

$$L_{D1} = -E_{x\sim G1(z)}[\min(0, -1 + D1(x)] - E_{x\sim G2(z)}[\min(0, -1 - D1(x)], \qquad (1)$$

$$L_{G1} = -E_{z\sim N}[\min(0, -1 + D1(G1(z))] - E_{z\sim N}[\min(0, -1 + D2(G1(z))], \qquad (2)$$

The loss function of Agent 2 is identical apart from exchanging D1/D2 and G1/G2. When this function is minimized, the discriminator learns to output positive and negative values for the patterns created by its own generator and opponent generator, respectively, and the generator is trained such that both discriminators output positive values. The batch size was set to 32, and one epoch represents a single parameter update by the losses computed using the 32 patterns produced by each generator.

To show that mutual adversarial imitation learning produces structural patterns, we compared its results to the unidirectional condition, where only one individual performs adversarial imitation learning. That model is denoted by blue arrows in Fig. 1. Under unidirectional adversarial imitation learning, there is no D2 discriminator, and the term D2 in Eq. (2) for $L_{G1}$ and $L_{G2}$ is omitted from the loss function.

## 3. Results

In our simulations, the model was trained for up to 5,000 epochs under both conditions. Figs. 2 and 3 show the patterns produces by each generator with eight different $z$ values under the unidirectional and mutual conditions, respectively. Because all the network parameters were initialized with random values, the generated patterns became equally cluttered at the beginning for both conditions.

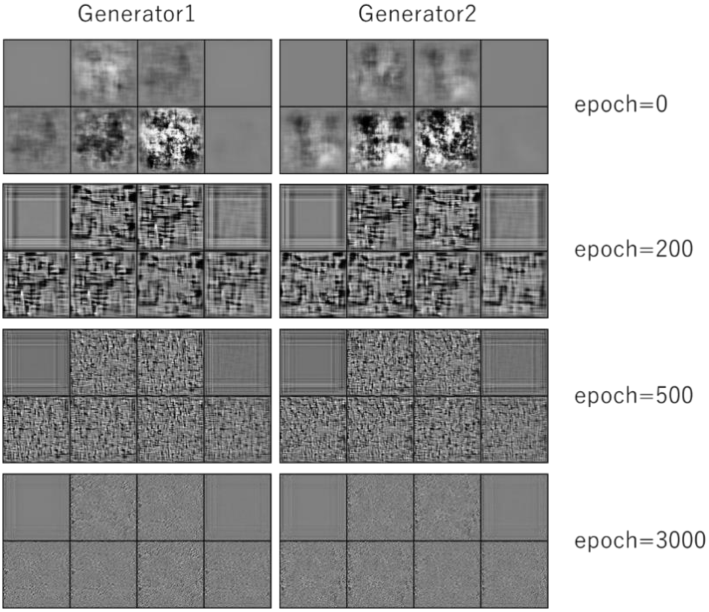


Figure 2. Examples of spatial patterns generated by G1 and G2 under unidirectional adversarial imitation learning. Each block has 8 patterns generated from 8 different $z$ values.

However, as the learning process progressed, the generated patterns became increasingly locally detailed and cluttered in unidirectional adversarial imitation learning. Under mutual adversarial imitation learning, however, these patterns appear to have formed a sort of global pattern.

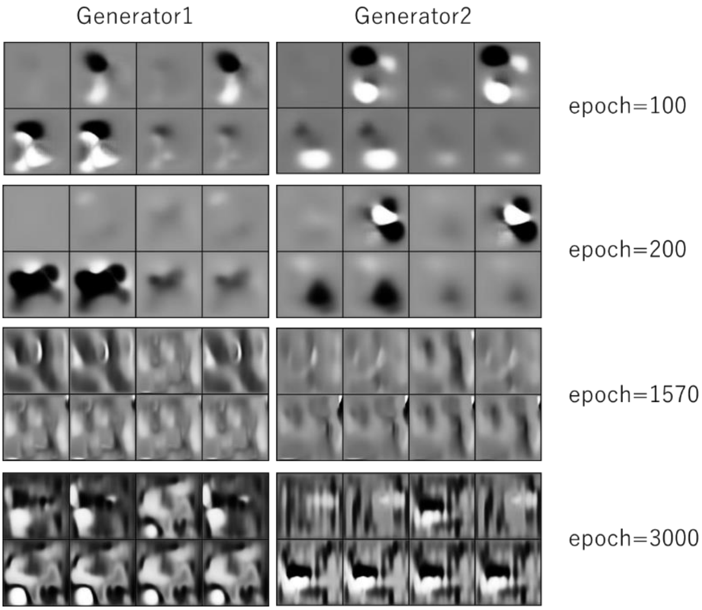


Figure 3. Examples of spatial patterns generated by G1 and G2 under mutual adversarial imitation learning.

To demonstrate the structuring progress under mutual adversarial imitation learning, we calculated the fractal dimensions of the generated patterns using the box-counting method with the patterns’ binary representations and the grayscale median. We also calculated the entropy of the grayscale images to measure complexity. Figure 4 shows the fractal dimensions and entropies for mutual and unidirectional adversarial imitation learning. Here, entropy decreased in the early stages of learning but remained high thereafter. The fractal dimension also did not increase, instead converging to a low value. As can be seen in Fig. 2, as the learning progresses, the pattern becomes more cluttered, but lacks an underlying structure.

Under mutual adversarial imitation learning, however, the entropy and fractal dimension exhibited oscillation. The entropy decreased, and the pattern became less cluttered, but the fractal dimension temporarily increased. Fig. 3 shows that the pattern was not merely cluttered but developed a structure. In the current simulation, this structure was not maintained, and the fractal dimension remained high. Nevertheless, the fractal dimension increased again, and structuring occasionally occurred. These results were confirmed through multiple simulations. Thus, our results show that mutual learning promotes the structuring of generated patterns.

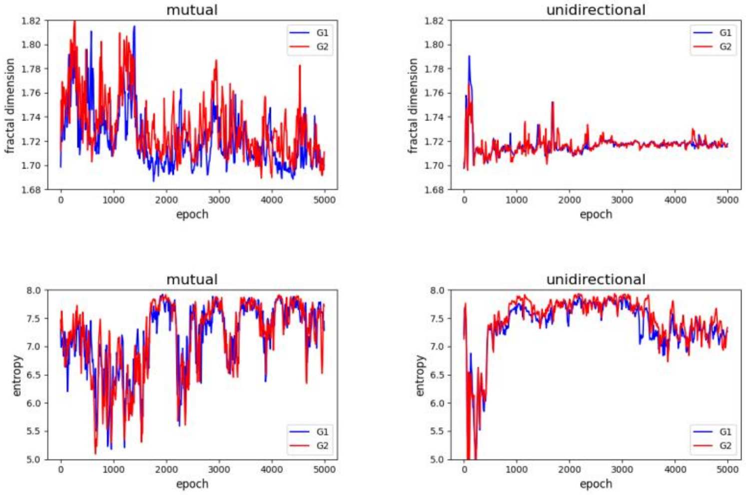


Figure 4. Changes in fractal dimensions and entropy of spatial patterns generated by G1 and G2. The left and right columns indicate learning results under the mutual and unidirectional conditions, respectively.

## 4. Discussion

In the case of unidirectional adversarial imitation learning, the discriminator/generator pair of Agent 1 should learn to escape from patterns created by the opponent generator. It is sufficient to gradually make the patterns chaotic to prevent the opponent from successfully imitating them. Once complexity arises, structuring can occur as a result of simplifying pressures, so that the patterns can be easily recognized by one's own discriminator. This corresponds to Kirby's group experiment of structuring sequences to be easily remembered in the process of cultural transmission through the learning force. In our previous model, the generated patterns became chaotic; however, simplification did not occur, and no structured patterns were found (Yamazaki et al., 2020). In the present model, the introduction of a discriminator makes it necessary to form patterns that are easy for the discriminator to recognize, which leads to pattern structuring. Because it is not sufficient for the generator to simply keep avoiding the opponent generator in mutual adversarial imitation learning, a stronger adversarial learning pressure was applied to the generators, resulting in structured patterns.

Our simulation results generated a high fractal dimension, indicating that the generated patterns had a recursive structure, similar to that of language. We believe that not only cooperative communication, but also adversarial communication was necessary to create such structures, which may have been incorporated into language.

## Acknowledgements

This work was supported by JSPS KAKENHI (grant number JP20H04989).

## References

Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution,* 2(2), 160–176.

Cholewiak, D. M., Sousa-Lima, R. S., & Cerchio, S. (2013). Humpback whale song hierarchical structure: Historical context and discussion of current classification issues. *Marine Mammal Science*, 29(3), E312-E332.

Chomsky N. (1965) Aspects of Theory of Syntax. Cambridge (MA): The MIT Press

Cornish, H., Smith, K., & Kirby, S. (2013). Systems from sequences: An iterated learning account of the emergence of systematic structure in a non-linguistic task. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35(35).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.

Hihara, S., Obayashi, S., Tanaka, M. and Iriki, A. (2003). Rapid learning of sequential tool use by macaque monkeys, *Physiology and Behavior*, 78, 427-434.

Honda, E., & Okanoya, K. (1999). Acoustical and syntactical comparisons between songs of the white-backed munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica). *Zoological Science*, 16(2), 319–326.

Jansen, D. A., Cant, M. A., & Manser, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (Mungos mungo) close calls. *BMC biology*, 10(1), 1-11.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686.

Kondo, S., & Miura, T. (2010). Reaction-diffusion model as a framework for understanding biological pattern formation. *Science*, 329(5999), 1616-1620.

Suzuki, J. and Kaneko, K. (1994). Imitation games. *Physica D: Nonlinear Phenomena*, 75(1–3):328–342.

Tokimoto, N., & Okanoya, K. (2004). Spontaneous construction of "Chinese boxes" by Degus (Octodon degu): A rudiment of recursive intelligence? *Japanese Psychological Research*, 46(3), 255-261.

Winters J., Kirby S., Smith K. (2015). Languages adapt to their contextual niche, *Language and Cognition*, 7(03): 415–49.

Yamazaki, S., Iizuka, H., & Yamamoto, M. (2020). Complexity of bird song caused by adversarial imitation learning. *Artificial Life and Robotics*, 25(1), 124–132.

# DEAF AND HARD-OF-HEARING PEOPLE CAN DETECT SOUND SYMBOLISM: IMPLICATIONS FOR THE ARTICULATORY ORIGIN OF WORD MEANING

Mutsumi Imai*[1], Kimi Akita[2], Sotaro Kita[3], Noburo Saji[4], Masato Ohba[5], and Miki Namatame[6]

*Corresponding Author: imai@sfc.keio.ac.jp
[1] Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan
[2] Department of English Linguistics, Nagoya University, Nagoya, Japan
[3]Department of Psychology, University of Warwick, Warwick, United Kingdom
[4] Faculty of Human Sciences, Waseda University, Tokorozawa, Japan
[5] Graduate School of Media and Governance, Keio University, Fujisawa, Japan
[6] Department of Industrial Information, Tsukuba University of Technology, Tsukuba, Japan

Uncovering the relation between form and meaning of words is important for understanding the nature and evolution of language. Where does the sound-meaning relationship of a word come from? In a long tradition of linguistics, it has been assumed that language-specific conventions create arbitrary association between linguistic forms and meanings (Hockett, 1960; de Saussure, 1916). However, words that sound like what they mean have also been recognized, and this intrinsic link is called sound symbolism (Köhler, 1929). Does one need to be able to hear sound to detect sound symbolism? Here, we show that one does not need to. Deaf-and-Hard-of-Hearing (DHH) participants, even those with profound hearing loss, could judge the sound symbolic match between shapes and words just as well as hearing participants do, as long as DHH participants could move their tongue freely. This indicates that people can detect sound symbolism via inherent resemblance between articulatory movements and the meanings. This further suggests that linguistic symbols can emerge through iconic mapping between oral gesture and sensory experience of the world.

## 1. Introduction

Uncovering the relation between form and meaning of words is important for understanding the nature and evolution of language. Form and meaning of words are associated not only by language-specific arbitrary conventions (de Saussure, 1916), but also by an intrinsic link, called sound symbolism (Köhler, 1929), which reflects perceived resemblance between sound and properties of referents. One may assume that the hearing ability is indispensable to sense sound symbolism. However, if sound symbolism is originated from articulatory gesture—mimicry of things or events by mouth—people may be able to detect sound-meaning match without the function of hearing. We propose *the articulatory origin of sound symbolism hypothesis*. We examined this hypothesis by testing whether Deaf or Hard of Hearing (DHH) people can detect sound symbolism.

## 2. Experiment 1.

### *2.1. Participants and Stimuli*

Thirty-four DHH individuals, all with congenital and profound hearing loss (minimum dB = 103), and 36 individuals with typical hearing participated in the study. Participants received a 2-page booklet. On each page, either a spiky-shaped figure or a round-shaped figure was depicted at the top, below which 38 novel words were listed orthographically. A total of 38 novel words were created. Previous findings (Köhler, 1929) indicate that, among these segments, sonorant consonants (/m, n, j, ɾ/), bilabials consonants (/b, p/), and non-front vowels (/a, o, ɯ/) are likely to be associated with round shapes, whereas non-labial obstruents (/d, t, g, k, z, s, ʑ, ɕ, tɕ, dʑ/) and front vowels (/i, e/) are good candidates for spiky shape sound symbolism (D’Onofrio, 2014; Nielsen & Rendall, 2011). These consonants and vowels were combined to form words with the /CVCV/ structure. The participants were instructed to judge whether each word matched the figure in three degrees: *good match*, *neutral*, and *mismatch*.

### *2.2. Results*

For each word-shape combination, we calculated the proportion of participants in each group who gave “good match” judgment (Figure 1). In Figure 1, the words were shown from the highest to the lowest acceptance (i.e., judgements of *good match*) by the hearing group for the round and spiky shapes. The distribution of match scores across items (word-shape combinations) was highly similar between the two groups: by-item correlations between the two groups (N = 38) were very high: $r = .893$ for the round figure (Figure 1a) and $r = .832$ for the spiky figure (Figure 1b). When both figures combined, $r = .862$, all $p$s < .001 (Figure 2). We split the DHH participants into two groups according to the severity of hearing loss to see whether DHH people’s judgements would differ depending on their hearing level. Even the severest group showed high correlations, which did not statistically differ from those shown by the milder group.

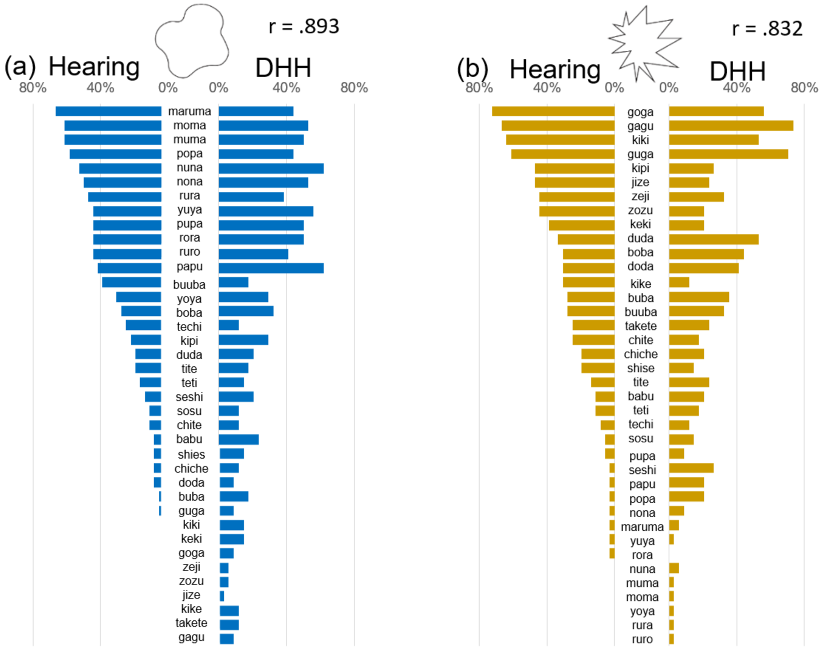


Figure 1. The percentage of participants in the hearing group and those in the DHH (Deaf and Hard of Hearing) group who judged the novel words to be good sound symbolic match with the shapes, in Experiment 1. The words are arranged in the descending order of the percentages for the hearing group.

A closer examination of Figures 1(a)(b) indicates that words that were judged good matching for the round shape tend to start with sonorant sounds, and those judged good matching for the spiky shape tend to start with obstruent sounds.

In Figure 2, the 38 words are plotted along the proportion of good-match judgements by the hearing (Y Axis) and the DHH group (X Axis). Here, the words starting with sonorant sounds are represented in green, and those starting with obstruent words are shown in red; values for the round shape and the spiky shape are represented in a circle and a triangle, respectively. Here, it is clear that the sonorant-round shape and obstruent-spiky shape combinations received high values, while the sonorant-spiky shape and obstruent-round shape combinations received low values in both groups.

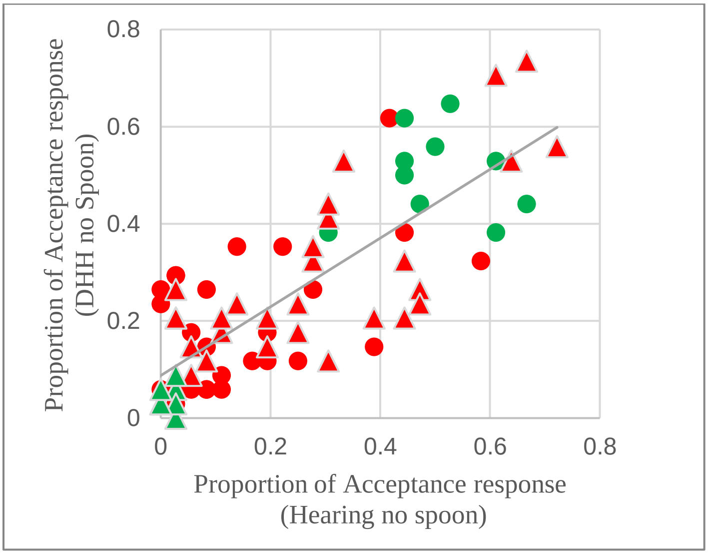


Figure 2. Plots of the 38 items along the proportion of good match judgements by the hearing group (X Axis) and by the DHH group (Y Axis).

### *2.3. Discussion*

The results supported *the articulatory origin of sound symbolism hypothesis* in that DHH people show judgements of sound-shape correspondences that are highly similar to those by hearing people. In both groups, participants judged sonorant-round shape and obstruent-spiky shape to be good matches, while the reverse combinations to be poor matches.

## 3. Experiment 2.

Experiment 2 examined whether disturbance of articulatory movements changes the judgements of sound-meaning correspondence both for hearing and DHH participants. In Experiment 2, participants placed a spoon on the tongue and closed the lips to hold the spoon, hence movement within the oral cavity and lips was generally restricted.

### *3.1. Participants and Stimuli*

Sixteen DHH students, all with congenital hearing loss, and 61 university students with typical hearing, participated. None of the participants took part in Experiment 1. The stimuli and the procedure were identical to Experiment 1, except that the participants were asked to hold a spoon with their mouth: they put the spoon in the oral cavity with the concave part of the spoon facing down, and placed the concave part on the tongue, and closed the lips to stabilize the handle of the spoon.

### *3.2. Results*

#### *3.2.1. By-item correlation analyses*

As in Experiment 1, a by-item correlation analysis was conducted. For the DHH group, the correlation coefficient between the baseline case (no-spoon in the hearing group, Experiment 1) and the *spoon on-tongue* case (Experiment 2) was $r = .639$, $p < .001$.This value was significantly lower than that between the two groups in the no-spoon case in Experiment 1. For the hearing group, the correlation coefficient between the baseline and the *on-tongue* case was still very high ($r = .913$, $p < .001$). The correlation coefficient in the *on-tongue* condition was significantly lower in the DHH group than in the hearing group (Meng's $z = 5.71$, $p = .000$). The correlation analyses thus endorsed the view that oral movement plays a critical role in sensitivity of sound symbolism.

#### *3.2.2. Model analysis*

The results of Experiment 1 suggested that the sonorant sounds and the obstruent sounds play a prominent role for the shape sound symbolism. Figure 3 shows the proportion of acceptance of words as good-matching in each of the sound/condition/group combinations for each shape. A mixed-effect logistic regression analysis was conducted to predict the participants' judgement (1:*good-match*/0: other responses), specifying Group (DHH:-0.5/Hearing: 0.5), Condition (articulation free:-0.5/articulation disturbed:0.5), shape (spiky:-0.5/round:0.5), sound (sonorant:-0.5/obstruent:0.5), and their interactions as fixed-effects. The participants are specified as random effects.

From the pattern shown in Figure 3, the interactions involving Sound x Shape strongly predict the proportion of good-match judgements.

The Shape x Sound interaction was highly significant, $z = -21.32$, $p = .000$. Also significant was two 3-way interactions: Condition x Shape x Sound interaction, $z = 3.08$, $p = .002$, and Group x Shape x Sound interaction, $z = -3.74$, $p = .000$. The effect of Condition was modified by Shape x Sound interaction in such a way that disturbance of articulation decreased the sense of good match for the round-sonorant combination but not so much for the obstruent-spiky combination. Likewise, the effect of Group was modified by the Shape x Sound interaction: the decrease of acceptance in round-sonorant combination was larger for the DHH group than for the Hearing group. However, the 4-way interaction effect was not significant.

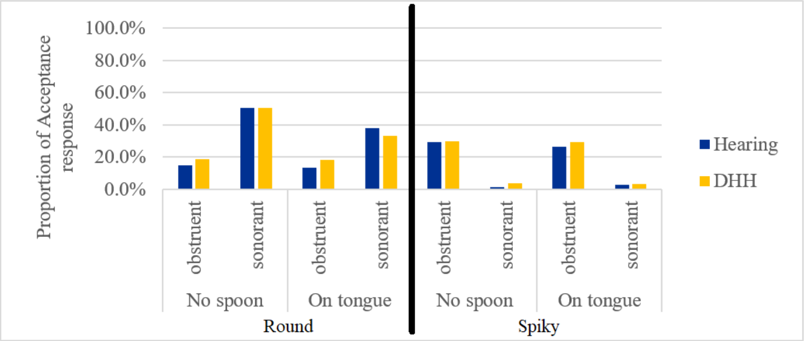


Figure 3. The proportion of acceptance of words as good-matching in each sound/condition/group combination for the Round shape (right panel) and for the Spiky shape (left panel).

### *3.3. Discussion*

The results strengthened the articulation origin of sound symbolism theory in that disturbance of articulatory movements affected sensitivity to sound symbolism. DHH individuals were affected by the disturbance of articulation more strongly than individuals with typical hearing and that the decrease of sensitivity to sound-shape correspondence mostly came from the sonorant sound-round shape correspondence, but not so much from the obstruent sound-spiky shape correspondence. The last result suggests that the mechanism of sound symbolism sensitivity is complex and that different mechanism may underlie for round shapes and spiky shapes (cf., Fort et al., 2018; Yang et al., 2019).

## 4. General Discussion

The results of the two experiments supported *the articulatory origin of sound symbolism hypothesis.* Our finding offers great insights into a key question in the literature of sound symbolism, i.e., whether sound symbolism arises from articulation (Sapir, 1929; Ramachandran & Hubbard, 2001; Margiotoudi & Pulvermüller, 2020). It is remarkable that Deaf and Hard-of-Hearing individuals can sense sound symbolism by mapping oral movements to the meanings. This strongly suggests that oral movements are crucial for sound symbolism detection. Furthermore, this fact highlights the view that sound symbolism is truly a multi-modal phenomenon, and not just the correspondence between audition and other sensory modalities (Perniss & Vigliocco, 2014; Pearlman et al., 2021). However, the current data do not rule out the possibility that hearing people recruit the acoustic information in addition to articulatory movements.

The current results have profound implications for theories of language evolution. Hand gesture is a prime candidate for how an open-ended shared lexicon emerged in early stages of language evolution (Stokoe, 2002; Arbib, 2005) because people can move their hands in a way that can iconically map to entities in the world such as events and objects(Goldin-Meadow et al., 1996; Ortega & Özyürek, 2020). The hand-gesture origin theories of language evolution maintain that hand movements have unique advantage over the movement of articulatory organs when it comes to iconically representing entities in the world(Stokoe, 2002; Arbib, 2005). However, the current result questions such assumption, and suggests that both hand movements and articulatory movements can be the basis of iconic meaning, offering a possible account for why language evolved in speech modality as well as in manual modality.

The current study has some limitations. First, because we presented the novel words orthographically, the possibility that DHH people's sensitivity may come from resemblance between the visual shape and shapes of letters. Although it is difficult to completely rule out the influence of letter symbolism, the results of Experiment 2 convincingly demonstrated that articulatory movements are critical for detecting sound-meaning correspondences (also see Cwiek et al, 2022 for the view that letter symbolism is minimum). Another limitation is that the spoon manipulation in Experiment 2 affected only the round shape-sonorant sound. Although this result suggests that different mechanism may underlie sound symbolism for round shapes and spiky shapes (cf. Fort et al., 2018, Yang et al., 2019), we cannot draw a conclusion concerning this issue, as our stimuli were not designed to address this issue.

## Acknowledgements

This study was supported by MEXT/JSPS KAKENHI Grant Numbers 20H00014, 16H01928, 18H05084 to Mutsumi Imai. We would like to thank Junko Kimura, Daiki Kondo, and Fumika Murayama for their help with data collection. We would also like to thank all participants in this study.

## References

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*, 105–124.

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., ... & Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B, 377,* 20200390.

D’Onofrio, A. (2014). Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Bouba-Kiki Paradigm. *Language and Speech, 57,* 367–393.

Fort, M., Lammertink, I., Peperkamp, S., Guevara ‐ Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: a meta ‐ analysis on the emergence of sound symbolism in early language acquisition. *Developmental science, 21,* e12659.

Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review, 103,* 34–55.

Köhler, W. (1929). *Gestalt Psychology. An Introduction to New Concepts in Modern Psychology*. Liveright Publishing Corporation.

Margiotoudi, K., & Pulvermüller, F. (2020). Action sound–shape congruencies explain sound symbolism. *Scientific Reports, 10,* 12706.

Ortega, G., & Özyürek, A. (2020). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods, 52,* 51–67.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130300.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia - A window into perception, thought and language. *Journal of Consciousness Studies, 8,* 3–34.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12,* 225-239.

de Saussure, F. (1916). *Course in General Lingusitics*. Columbia University Press.

Stokoe, W. (2002). *Language in Hand: Why Sign Came Before Speech*. Gallaudet University Press.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological bulletin, 111,* 172-175.

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology, 65,* 115–124.

Yang J, Asano M, Kanazawa S, Yamaguchi MK, Imai M. (2019). Sound symbolism processing is lateralized to the right temporal region in the prelinguistic infant brain. *Scientific Report, 9,* 13435.

# MALE NORTHERN GRAY GIBBON SONGS HAVE A PHRASE-INSERTING STRUCTURE

YOICHI INOUE*[1], WAIDI SINUN[2], SHIGETO YOSIDA[3] and KAZUO OKANOYA[4]

*Corresponding Author: gibbonyoichi@hotmail.co.jp
[1]Cognition and Behavior Joint Research Laboratory, Riken Center for Brain Science Institute, Wako, Japan
[2]Research and Development Division Yayasan Sabah Group, Kota Kinabalu, Malaysia
[3]IQVIA Solusions Japan K.K., Tokyo, Japan
[4]Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan

## 1. Introduction

To date, there have been no reports of a phrase-embedding structure in animal vocal communication, although there are several reports of combinational animal sounds (Hauser et al., 2014). Gibbons (Hylobatidae) living in South-East Asia are small apes, among the closest living relatives of humans. Gibbons are known for their remarkable vocal behavior. The songs of male Northern Gray gibbons (*Hylobates funereus*) in the Danum Valley Conservation Area (DVCA), Sabah, Malaysia consist of two notes, wa" and "oo", and combinations of these are flexible (Inoue et al., 2017). The "wa" notes were generally short, with a rapid rise in frequency, and the "oo" notes were relatively monotonal. The phrases of a song have various note orders, i.e., the notes could be combined in different phrase types. The Northern Gray gibbon is the species formerly named as Mueller's gibbon (*Hylobates muelleri*).

When phrase "N" is inserted within another phrase "AB", the generated phrase is shown as "ANB". We named this structure a phrase-inserting structure. In male Northern Gray gibbon songs in our study area, a succession of "wa" notes (trill) are sometimes placed at both the start and end of phrases. When trills were placed at both the start and end of AB phrases, we named these AB phrases fixed phrases. Trills were located in front of "A" and at end of "B." In this case, each A and B has a pairwise relationship and always cooccurs in a string. In the middle of a fixed phrase "AB," notes consisting of three or more notes were sometimes included. If the included notes were also sung as a phrase independently, we defined it as “N.”

If the phrase “N” itself also has an inserting structure, then ANB is represented as AABB, which can be called an embedding structure, as defined by Abe and Watanabe (2011). When the animals sing songs by combining acoustic elements flexibly, a complicated syntax may emerge. We hypothesized that there are phrase-inserting structures in male Northern Gray gibbon songs, and recorded songs by a male in the DVCA for analysis. This report is based on a paper by Inoue et al. (2020).

## 2. Methods

All observations were conducted in the DVCA located in Sabah, Malaysia, in the northeast region of Borneo Island (Fig. 1). We studied a gibbon group named “SAPA”, whose territory was located around the Borneo Rainforest Lodge (BRL). With the aid of a field assistant, we conducted a survey over 4–7 successive days, biannually in August and December, from 2001 to 2009. During the study period from 2001 to 2009, we recorded 70 songs in 107 days of observations. We arrived the sleeping trees where the SAPA group slept on the previous day. We started following at 0500 h and ended about 30 min after the gibbons had arrived at their sleeping trees. We started recording gibbon songs, as soon as they started singing. We used a digital audio recorder (R-09; Roland, Hamamatsu, Japan) with a microphone (ATM57; Audio-Technica, Tokyo, Japan), and recorded the gibbon voices under the trees in which the study male was singing. The recorder was set at a 44.1-kHz sampling rate and had 16-bit resolution. We analyzed 8,046 phrases in 70 songs from the SAPA adult male. We converted the recorded sounds to sonograms using Avisoft-SAS Lab Pro software (Avisoft, Berlin, Germany). Focusing on the fundamental frequency, we performed the spectral and temporal measurements described below. To remove ambient noise, we processed the sound through a high-pass filter to cut off sound below 500 Hz. Finally, sonograms were created for on-screen measurements (settings: 256-point fast Fourier transformation and Hamming windows). A song is a series of notes, generally of more than one type, uttered in succession and so related as to form a recognizable sequence or pattern in time (Thorpe, 1961). A phrase is a larger, loose collection of several notes preferentially voiced in combination. Intra-phrase intervals are shorter than inter-phrase intervals (Geissmann et al., 2005). Most of the note intervals in our subject male’s songs were less than 2.0 seconds (90.7%). Therefore, we identified different phrases within a song as being separated by pauses of > 2 seconds.

In order to confirm two note types, we selected a 5 % random sample of all notes and measured the onset time, offset time, start frequency, end frequency, middle frequency, maximum frequency, and minimum frequency of each note and inter-note intervals for each song phrase. Based on the measured data, we calculated the duration, Δ frequency, and Δ frequency/duration of each note. Two-tailed t tests were used to compare the acoustic characteristics of the two notes.

Furthermore, we analyzed whether or not the corresponding phrases were indistinguishable in terms of the acoustic similarity between those with an inserted phrase (A-N-B) and with no inserted phrase (A-B). We selected 20 patterns of phrases and calculated the acoustical similarity using Avisoft-CORRELATOR for every pair of same type phrases. We calculated the acoustical similarity between all pairs of "A" and "B" notes within the A-N-B phrases (within-class condition), within the A-B phrases (within-class condition), and between the A-N-B and the A-B phrases (between-class condition). The Avisoft-CORRELATOR allowed us to compute cross-correlations between spectrograms by sliding them along with the time axis. The approach of the Avisoft-CORRELATOR can be compared to computing correlations between two grayscale raster images, while sliding them on the X-axis. The highest correlation coefficients for each pair of sounds were regarded as similarity scores. We examined whether the acoustical similarity was the same regardless of whether the pair of phrases were both from the same class (within-class condition) or from different classes (between-class condition). We performed a linear mixed-effect model entering the cross-correlation coefficients as the response variable and the within/between-class condition as the explanatory variable (fixed-effect). We entered the phrase types (the above-mentioned 20 patterns of phrases) as random intercepts because the average similarity scores were expected to differ among them. We performed model diagnosis by visually assessing normality and homogeneity of residual variance across the random groups and normality of the random intercepts using group-wise boxplots and normal quantile-quantile plots. We reported the mean and standard deviations of the similarity score for the within/between-class condition, and examined the statistical significance using a likelihood-ratio test (Pinheiro and Bates, 2000). R 3.6.1 (R Core Team 2019) and nlme package (Pinheiro et al., 2019) were used for the analysis.

## 3. Results

The mean values of the all examined acoustic parameters of notes were statistically different between the two notes (Table 1). In this study, three or more successive "wa" notes (trill) were placed at both the start and end of phrases in

2,726 (33.9%) of the 8,046 phrases. Among these, we identified 1,891 phrases (23.5%) as fixed phrases. Phrases consisting of three or more notes were sometimes included within fixed phrases (Fig. 2). We found 448 phrases (5.6%) that included phrases consisting of three or more notes.

The overall mean similarities for the within-class and between-class conditions were 0.52 (SD = 0.12, n =1290) and 0.50 (SD = 0.11, n = 1346), respectively. The diagnostic plots of the linear mixed effects model showed no indication of violating the model assumption (result not shown). The result of the linear mixed effects model showed that the similarity was estimated to be 0.02 (standard error = 0.003) higher in the within-class condition, after controlling the phrase type (X1 = 33.15, $p < 0.01$).

## 4. Discussion

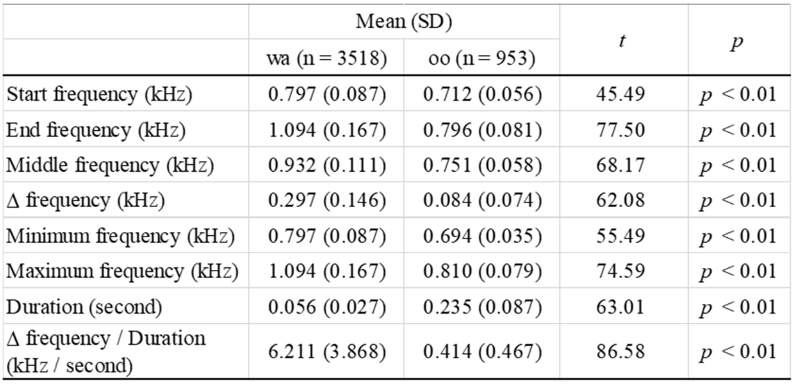
We analyzed songs by a single male. There were no notes in the studied gibbon's songs with acoustic characteristics largely different from "wa" and "oo" notes. The male had various types of note orders in his phrases. In his songs, we identified fixed phrases, which were characterized by three or more successive "wa" notes (trills) placed at both the start and end of phrases. Phrases consisting of three or more notes were sometimes included within fixed phrases. We found 448 phrases (5.6%) that included phrases consisting of three or more notes. We have concluded that these note orders suggest a phrase-inserting structure.

Corresponding phrases with an inserting structure and no inserting structure were very similar in terms of sonograms, and hardly discriminable by human ear. However, the statistical analysis showed a small but systematic difference in acoustic similarity between them; phrases were more similar among pairs that both had an inserting structure or did not have an inserting structure (within-class condition), than among pairs where one had an inserting structure and the other did not (between-class condition). Further study under experimental conditions will be needed to examine whether gibbons are able to discriminate this small acoustic difference.

One of the main differences between language and non-human animal communication is the grammar used to produce sequences. Human language uses "context-free grammars" that are capable of generating recursive sequences (Chomsky, 2002). Among various discussions about the definition of recursion (e.g., Fitch, 2010), there is an interpretation that recursion consists of embedding a constituent into a constituent of the same type (e.g., Pinker and Jackendoff, 2005; Martins and Fitch, 2014). In contrast, animal vocal sequences are usually described as "regular grammars", a simple kind of concatenation system (Berwick et al., 2012). Many researchers have considered that non-human animal vocalizations would belong to regular ones. This paper tested the grammatical structures of primate "songs" with the comparative perspectives of the human language syntax and that of other animals. Our results showed that male gibbon's songs have a phrase-inserting structure which is considered to be a precursory level to recursion.

A phrase-inserting structure occurred when combinations of notes in the male's songs were flexible. Our results lead us to hypothesize that complicated syntax emerges when the animals sing songs by combining acoustic elements flexibly. This may be the first evidence of a phrase-inserting structure in animal songs. Our data and linguistic perspectives may certainly be of use in future studies to elucidate vocal communication in gibbons and other non-human primates. However, as we collected data from only one male, further studies on many gibbon groups will be necessary to confirm our results.

Table 1. Acoustic characteristics of "wa" and "oo" notes.



| | Mean (SD) | | *t* | *p* |
|---|---|---|---|---|
| | wa (n = 3518) | oo (n = 953) | | |
| Start frequency (kHz) | 0.797 (0.087) | 0.712 (0.056) | 45.49 | $p < 0.01$ |
| End frequency (kHz) | 1.094 (0.167) | 0.796 (0.081) | 77.50 | $p < 0.01$ |
| Middle frequency (kHz) | 0.932 (0.111) | 0.751 (0.058) | 68.17 | $p < 0.01$ |
| Δ frequency (kHz) | 0.297 (0.146) | 0.084 (0.074) | 62.08 | $p < 0.01$ |
| Minimum frequency (kHz) | 0.797 (0.087) | 0.694 (0.035) | 55.49 | $p < 0.01$ |
| Maximum frequency (kHz) | 1.094 (0.167) | 0.810 (0.079) | 74.59 | $p < 0.01$ |
| Duration (second) | 0.056 (0.027) | 0.235 (0.087) | 63.01 | $p < 0.01$ |
| Δ frequency / Duration (kHz / second) | 6.211 (3.868) | 0.414 (0.467) | 86.58 | $p < 0.01$ |

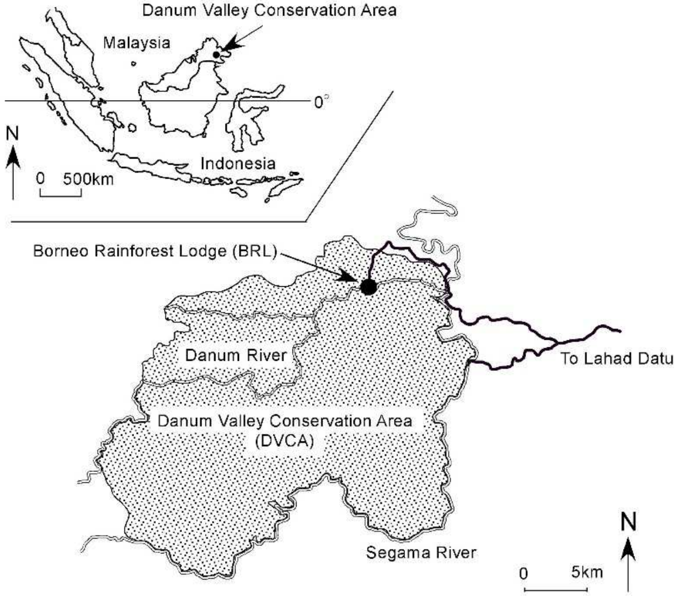


Figure 1. Location of the Borneo Rainforest Lodge (BRL) in the Danum Valley Conservation Area, Sabah, Malaysia.

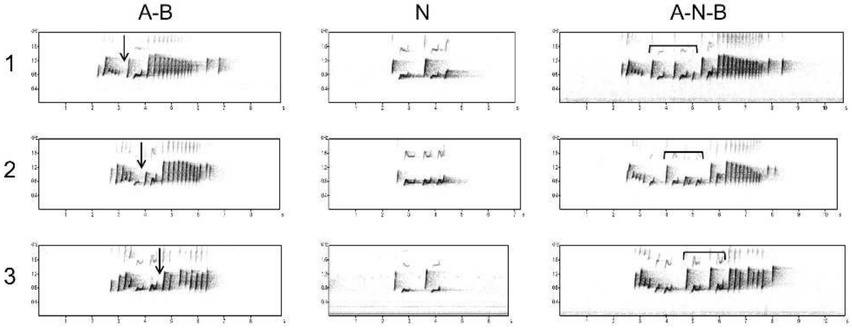


Figure 2. Examples of phrase A-N-B included within fixed phrases. Arrows indicate where phrase N was included within phrase A-B. Solid line brackets indicate phrase N included within phrase A-B.

## 5. References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience, 14*, 1067–1074.

Berwick, R. C., Beckers, G. J., Okanoya, K., & Bolhuis, J. J. (2012). A bird's eye view of human language evolution. *Frontiers in Evolutionary Neuroscience, 4,* 5.

Chomsky, N. (2002). Syntactic structures, 9th edn. The Hague, The Netherlands: de Gruyter Mouton.

Fitch, W. T. (2010). Three meanings of 'recursion': key distinctions for biolinguistics. In Larson, R. K., Deprez, V., & Yamakido, H. (eds): The evolution of human language: Biolinguistic perspectives, Cambridge: Cambridge University Press, pp. 73–90.

Geissmann, T., Bohlen-Eyring, S., & Heuck, A. (2005). The male song of the Javan silvery gibbon (*Hylobates moloch*). *Contributions to Zoology, 74*, 1–25.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N., & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in Psychology, 5*, 401.

Inoue, Y., Sinun, W., Yosida, S., & Okanoya, K. (2017). Combinatory rules and chunk structure in male Mueller's gibbon songs. *Interaction Studies 18*, 1–25.

Inoue, Y., Sinun, W., Yosida, S., & Okanoya, K. (2020). Note orders suggest phrase-inserting structure in male Mueller's gibbon songs: a case study. *acta ethologica*, 23(2), 89–102.

Martins, M. D., & Fitch, W. T. (2014). Investigating recursion within a domain-specific framework. In Language and recursion. Springer, New York, NY. pp. 15–26.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2019). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–143. URL: https://CRAN.R-project.org/package=nlme.

Pinheiro, J. C., & Bates, D. M. (2000). "Mixed-Effects Models in S and S-PLUS", New York: Springer.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition, 95(2),* 201–236.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Thorpe, W. H. (1961). Bird-song. The biology of vocal communication and expression in birds. Cambridge monographs in experimental biology No. 12. Cambridge: University Press.

# ELEMENTS FOR A CULTURAL EVOLUTION OF SUPERNATURAL RITUAL LANGUAGES

ARITZ IRURTZUN

aritz.irurtzun@iker.cnrs.fr
CNRS–IKER (UMR 5478), Bayonne, France

I propose a characterization of ritual languages as ‘natural conlangs’, comparable to the ‘laboratory conlangs’ of iterated learning experiments. I present the Ritual Language DataBase (RLDB), which provides a typology of such languages around the world. A comparative analysis of the RLDB uncovers patterns, and the significant role of users (community size), functions, and alternate languages when shaping the grammatical structure of ritual languages.

## 1. Introduction: Speakers and communities as actors in language evolution

Formal studies of language conceive of language acquisition and language change as a dynamic process (*cf.* Andersen, 1973; Lightfoot, 1979; Yang, 2002):

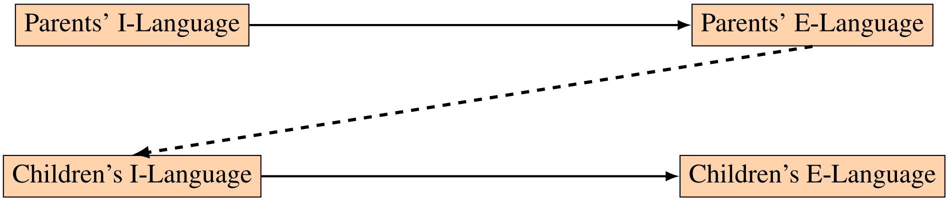


Figure 1. Dynamics of language acquisition and change (adapted from Andersen, 1973: 767).

The idea is that children (*i.e.*, the language acquirers) do not have direct access to the grammar/I(nternal)-language of their parents, but an indirect one, through the E(xternal)-language that they parse. This E-language will include some signatures or triggers that will guide the children in the choice of the ‘correct’ (compatible) target I-language. But rather than this being a deterministic system, language acquirers have an active role in shaping their language, as they have to (re)interpret the E-language of their environment (their primary linguistic data).

What is more, over the last decades intensive research on language evolution and change has uncovered some of the biases that humans act upon when acquiring their languages (*cf.* Kirby, 2017; Culbertson, Smolensky, & Legendre, 2012; Biberauer & Roberts, 2017; Raviv, Meyer, & Lev-Ari, 2019).

In this talk I want to extend this research paradigm to a new area: I give the first steps towards a characterization of the typology and analysis of the cultural evolution of the linguistic practices employed for supernatural ritual functions.

## 2. Supernatural rituals and 'natural' conlangs

Very often, in all sorts of cultural niches across the world, people change their language when performing supernatural ritual functions. I analyze these special languages as some sort of 'natural' conlangs, comparable to the 'laboratory' conlangs usually employed in iterated learning research on language evolution.

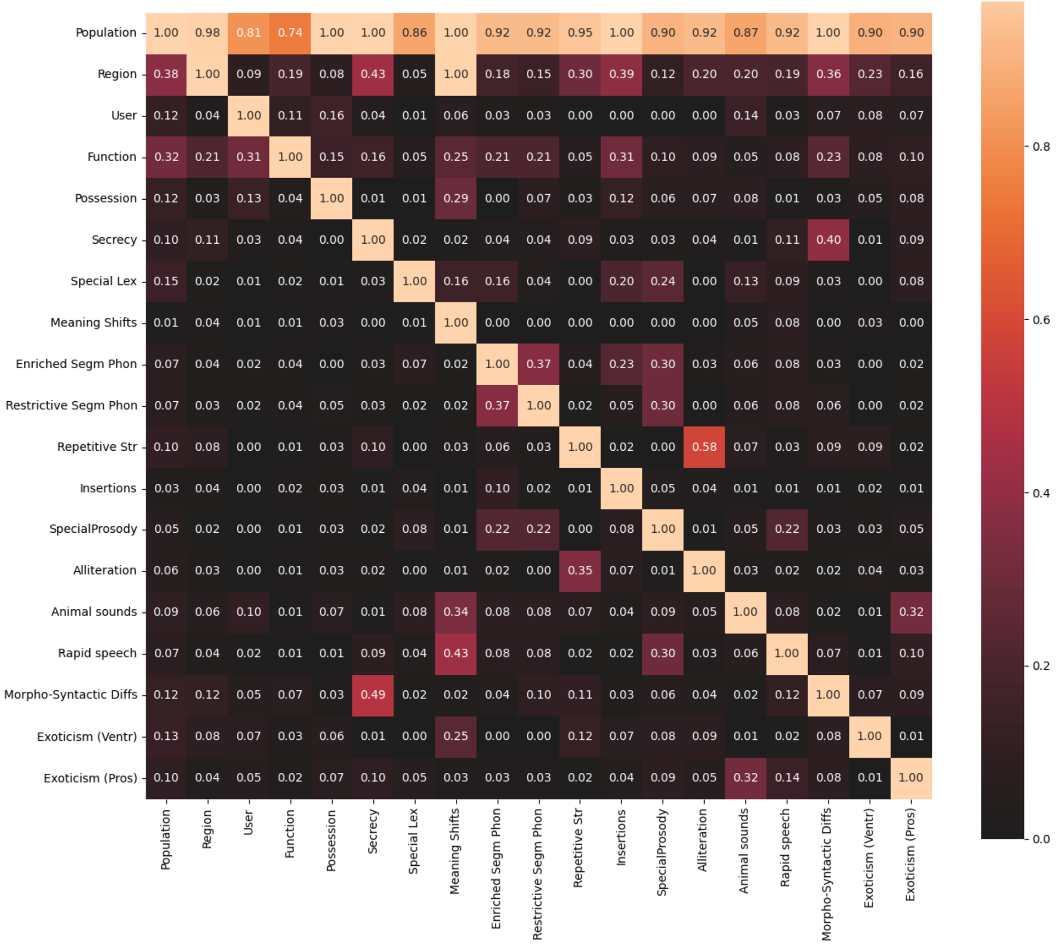
### 2.1. *The Ritual Language Database*

In order to analyze the nature and variability of ritual languages, I designed the Ritual Language Database (RLDB), which at the moment comprises data from 242 linguistic practices related to supernatural rituals across a wide range of different populations across the world.

The RLDB systematically documents 46 features of each ritual linguistic use such as the population (Amharas, Antakarinjas, Apaches...); its function (whether it is employed for cursing, divination, fortune-seeking, healing, initiation, mourning, or order-keeping); its user (whether it is a designated practitioner individual (a shaman, a priest, etc.) or a larger community that employs it); whether it implies possession by a spirit or ghost; whether it implies an alternate language (which can be oral or gestual); whether the alternate language is an archaic variant of the local language or an independent language altogether; whether it implies glossolalia; use of ventriloquism, animal and nature sounds, etc.; and if there is no alternate language, whether the language employed has a special lexicon (which can be constituted of archaisms, borrowings, or ad hoc lexical items); whether it comprises meaning shifts (metaphors, circumlocutions); whether its segmental inventory is enriched with respect to the normal language (a larger set of consonants/vowels, lessened phonotactic restrictions) or impoverished (a more restrictive set of consonants/vowels or stronger phonotactic constraints); whether it involves aliterations and syntagmatic repetitions; morpho-phonological insertions; a special prosodic pattern; morphological and syntactic differences, etc. The RLDB is a dynamic database, developed continuously with new data from ethnographies, linguistic descriptions and religion studies reports.

### 2.2. *Association between linguistic features*

The linguistic phenomena of supernatural rituals at the RLDB show different degrees of association. These are summarized in Figs. 2 and 3 below. Fig. 2 displays the association between features in ritual practices not involving an alternate language, whereas Fig. 3 is devoted to those practices that do involve an alternate language. Numbers report the uncertainty coefficient (Theil's U), a measure of conditional entropy between variables (ranging from 0 to 1, where there is no possible negative association). As such, this is an asymmetric measure, and thus we can obtain an estimate about the nature of variable B, given our knowledge of variable A, which is more informative than a simple (symmetric) correlation. For instance, Fig. 2 shows that knowing the function of the ritual can be more informative regarding the user ($U = .31$) than the other way around ($U = .11$).[1]

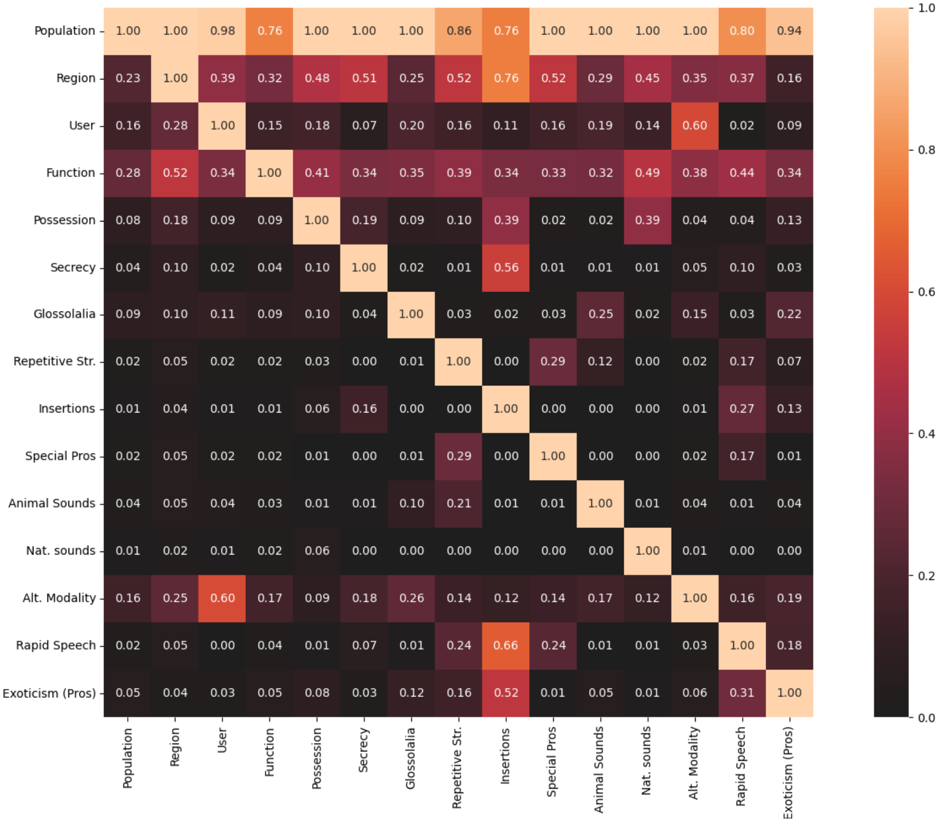


| | Population | Region | User | Function | Possession | Secrecy | Special Lex | Meaning Shifts | Enriched Segm Phon | Restrictive Segm Phon | Repetitive Str | Insertions | SpecialProsody | Alliteration | Animal sounds | Rapid speech | Morpho-Syntactic Diffs | Exoticism (Ventr) | Exoticism (Pros) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | 1.00 | 0.98 | 0.81 | 0.74 | 1.00 | 1.00 | 0.86 | 1.00 | 0.92 | 0.92 | 0.95 | 1.00 | 0.90 | 0.92 | 0.87 | 0.92 | 1.00 | 0.90 | 0.90 |
| Region | 0.38 | 1.00 | 0.09 | 0.19 | 0.08 | 0.43 | 0.05 | 1.00 | 0.18 | 0.15 | 0.30 | 0.39 | 0.12 | 0.20 | 0.20 | 0.19 | 0.36 | 0.23 | 0.16 |
| User | 0.12 | 0.04 | 1.00 | 0.11 | 0.16 | 0.04 | 0.01 | 0.06 | 0.03 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.03 | 0.07 | 0.08 | 0.07 |
| Function | 0.32 | 0.21 | 0.31 | 1.00 | 0.15 | 0.16 | 0.05 | 0.25 | 0.21 | 0.21 | 0.05 | 0.31 | 0.10 | 0.09 | 0.05 | 0.08 | 0.23 | 0.08 | 0.10 |
| Possession | 0.12 | 0.03 | 0.13 | 0.04 | 1.00 | 0.01 | 0.01 | 0.29 | 0.00 | 0.07 | 0.03 | 0.12 | 0.06 | 0.07 | 0.08 | 0.01 | 0.03 | 0.05 | 0.08 |
| Secrecy | 0.10 | 0.11 | 0.03 | 0.04 | 0.00 | 1.00 | 0.02 | 0.02 | 0.04 | 0.04 | 0.09 | 0.03 | 0.03 | 0.04 | 0.01 | 0.11 | 0.40 | 0.01 | 0.09 |
| Special Lex | 0.15 | 0.02 | 0.01 | 0.02 | 0.01 | 0.03 | 1.00 | 0.16 | 0.16 | 0.04 | 0.00 | 0.20 | 0.24 | 0.00 | 0.13 | 0.09 | 0.03 | 0.00 | 0.08 |
| Meaning Shifts | 0.01 | 0.04 | 0.01 | 0.01 | 0.03 | 0.00 | 0.01 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.08 | 0.00 | 0.03 | 0.00 |
| Enriched Segm Phon | 0.07 | 0.04 | 0.02 | 0.04 | 0.00 | 0.03 | 0.07 | 0.02 | 1.00 | 0.37 | 0.04 | 0.23 | 0.30 | 0.03 | 0.06 | 0.08 | 0.03 | 0.00 | 0.02 |
| Restrictive Segm Phon | 0.07 | 0.03 | 0.02 | 0.04 | 0.05 | 0.03 | 0.02 | 0.02 | 0.37 | 1.00 | 0.02 | 0.05 | 0.30 | 0.00 | 0.06 | 0.08 | 0.06 | 0.00 | 0.02 |
| Repetitive Str | 0.10 | 0.08 | 0.00 | 0.01 | 0.03 | 0.10 | 0.00 | 0.03 | 0.06 | 0.03 | 1.00 | 0.02 | 0.00 | 0.58 | 0.07 | 0.03 | 0.09 | 0.09 | 0.02 |
| Insertions | 0.03 | 0.04 | 0.00 | 0.02 | 0.03 | 0.01 | 0.04 | 0.01 | 0.10 | 0.02 | 0.01 | 1.00 | 0.05 | 0.04 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 |
| SpecialProsody | 0.05 | 0.02 | 0.00 | 0.01 | 0.03 | 0.02 | 0.08 | 0.01 | 0.22 | 0.22 | 0.00 | 0.08 | 1.00 | 0.01 | 0.05 | 0.22 | 0.03 | 0.03 | 0.05 |
| Alliteration | 0.06 | 0.03 | 0.00 | 0.01 | 0.03 | 0.02 | 0.00 | 0.01 | 0.02 | 0.00 | 0.35 | 0.07 | 0.01 | 1.00 | 0.03 | 0.02 | 0.02 | 0.04 | 0.03 |
| Animal sounds | 0.09 | 0.06 | 0.10 | 0.01 | 0.07 | 0.01 | 0.08 | 0.34 | 0.08 | 0.08 | 0.07 | 0.04 | 0.09 | 0.05 | 1.00 | 0.08 | 0.02 | 0.01 | 0.32 |
| Rapid speech | 0.07 | 0.04 | 0.02 | 0.01 | 0.01 | 0.09 | 0.04 | 0.43 | 0.08 | 0.08 | 0.02 | 0.02 | 0.30 | 0.03 | 0.06 | 1.00 | 0.07 | 0.01 | 0.10 |
| Morpho-Syntactic Diffs | 0.12 | 0.12 | 0.05 | 0.07 | 0.03 | 0.49 | 0.02 | 0.02 | 0.04 | 0.10 | 0.11 | 0.03 | 0.06 | 0.04 | 0.02 | 0.12 | 1.00 | 0.07 | 0.09 |
| Exoticism (Ventr) | 0.13 | 0.08 | 0.07 | 0.03 | 0.06 | 0.01 | 0.00 | 0.25 | 0.00 | 0.00 | 0.12 | 0.07 | 0.08 | 0.09 | 0.01 | 0.02 | 0.08 | 1.00 | 0.01 |
| Exoticism (Pros) | 0.10 | 0.04 | 0.05 | 0.02 | 0.07 | 0.10 | 0.05 | 0.03 | 0.03 | 0.03 | 0.02 | 0.04 | 0.09 | 0.05 | 0.32 | 0.14 | 0.08 | 0.01 | 1.00 |

Figure 2. Heatmap of the association between features in the RLDB for ritual practices that do not involve an alternate language.

---

[1] For comparison, the symmetrical measure Cramér's V associates the two variables at $V = .50$.



| | Population | Region | User | Function | Possession | Secrecy | Glossolalia | Repetitive Str. | Insertions | Special Pros | Animal Sounds | Nat. sounds | Alt. Modality | Rapid Speech | Exoticism (Pros) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Population | 1.00 | 1.00 | 0.98 | 0.76 | 1.00 | 1.00 | 1.00 | 0.86 | 0.76 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 0.94 |
| Region | 0.23 | 1.00 | 0.39 | 0.32 | 0.48 | 0.51 | 0.25 | 0.52 | 0.76 | 0.52 | 0.29 | 0.45 | 0.35 | 0.37 | 0.16 |
| User | 0.16 | 0.28 | 1.00 | 0.15 | 0.18 | 0.07 | 0.20 | 0.16 | 0.11 | 0.16 | 0.19 | 0.14 | 0.60 | 0.02 | 0.09 |
| Function | 0.28 | 0.52 | 0.34 | 1.00 | 0.41 | 0.34 | 0.35 | 0.39 | 0.34 | 0.33 | 0.32 | 0.49 | 0.38 | 0.44 | 0.34 |
| Possession | 0.08 | 0.18 | 0.09 | 0.09 | 1.00 | 0.19 | 0.09 | 0.10 | 0.39 | 0.02 | 0.02 | 0.39 | 0.04 | 0.04 | 0.13 |
| Secrecy | 0.04 | 0.10 | 0.02 | 0.04 | 0.10 | 1.00 | 0.02 | 0.01 | 0.56 | 0.01 | 0.01 | 0.01 | 0.05 | 0.10 | 0.03 |
| Glossolalia | 0.09 | 0.10 | 0.11 | 0.09 | 0.10 | 0.04 | 1.00 | 0.03 | 0.02 | 0.03 | 0.25 | 0.02 | 0.15 | 0.03 | 0.22 |
| Repetitive Str. | 0.02 | 0.05 | 0.02 | 0.02 | 0.03 | 0.00 | 0.01 | 1.00 | 0.00 | 0.29 | 0.12 | 0.00 | 0.02 | 0.17 | 0.07 |
| Insertions | 0.01 | 0.04 | 0.01 | 0.01 | 0.06 | 0.16 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.27 | 0.13 |
| Special Pros | 0.02 | 0.05 | 0.02 | 0.02 | 0.01 | 0.00 | 0.01 | 0.29 | 0.00 | 1.00 | 0.00 | 0.00 | 0.02 | 0.17 | 0.01 |
| Animal Sounds | 0.04 | 0.05 | 0.04 | 0.03 | 0.01 | 0.01 | 0.10 | 0.21 | 0.01 | 0.01 | 1.00 | 0.01 | 0.04 | 0.01 | 0.04 |
| Nat. sounds | 0.01 | 0.02 | 0.01 | 0.02 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.01 | 0.00 | 0.00 |
| Alt. Modality | 0.16 | 0.25 | 0.60 | 0.17 | 0.09 | 0.18 | 0.26 | 0.14 | 0.12 | 0.14 | 0.17 | 0.12 | 1.00 | 0.16 | 0.19 |
| Rapid Speech | 0.02 | 0.05 | 0.00 | 0.04 | 0.01 | 0.07 | 0.01 | 0.24 | 0.66 | 0.24 | 0.01 | 0.01 | 0.03 | 1.00 | 0.18 |
| Exoticism (Pros) | 0.05 | 0.04 | 0.03 | 0.05 | 0.08 | 0.03 | 0.12 | 0.16 | 0.52 | 0.01 | 0.05 | 0.01 | 0.06 | 0.31 | 1.00 |

Figure 3. Heatmap of the association between features in the RLDB for ritual practices that do involve an alternate language.

An overview of the data shows that the linguistic features of ritual languages are not distributed at random but tend to cluster in patterned ways, which suggests the effect of universal biases. This is a remarkable fact, given the fragmentary and uneven evidence reported in the ethnographic and linguistic records from populations around the world.[2]

Due to space restrictions here I can only comment a couple of them. Concentrating on function, we can observe that there are sharp differences *vis-à-vis* the employment of an alternate language ($X^2(6, N = 217) = 38.584; p < .001$):

---

[2] The RLDB comprises references as varied as the recent and scholarly Jorgensen (2020) on the sign language of the Balgo or Hall (2019) on the incantations of Hup shamans, as weel as ancient ethnographic reports such as Gabb (1875) on the Talamancans, or Martin (1817) on the Tongans.

Table 1. Function $\sim$ alt. language

| Function | Alternate language<br>not available | <br>available | Total |
|---|---|---|---|
| cursing | 1 | 3 | 4 |
| divination | 12 | 8 | 20 |
| fortune | 21 | 6 | 27 |
| healing | 25 | 18 | 43 |
| initiation | 25 | 30 | 55 |
| keep order | 10 | 5 | 15 |
| mourning | 8 | 45 | 53 |
| TOTAL | 102 | 115 | 217 |

For instance, it is more likely for people to employ an alternate language in mourning than in fortune-seeking rituals ($X^2(2, N = 80) = 8.2573; p < .004$). Not only that, within the set of practices that do employ an alternate language, use of an alternate modality (a sign language like Marumpu Wangka employed by the Balgo Aboriginals (Jorgensen, 2020)) is also non-randomly distributed across functions ($X^2(6, N = 107) = 45, 752; p < .001$):

Table 2. Function $\sim$ alt. modality in [+alt. langs]

| Function | Alternate modality<br>not available | <br>available | Total |
|---|---|---|---|
| cursing | 3 | 0 | 3 |
| divination | 7 | 0 | 7 |
| fortune | 5 | 1 | 6 |
| healing | 16 | 0 | 16 |
| initiation | 9 | 18 | 27 |
| keep order | 3 | 1 | 4 |
| mourning | 10 | 34 | 44 |
| TOTAL | 53 | 54 | 107 |

Last, in the populations that do not employ an alternate language but resort to a variant of their language (a natural conlang), we can observe sharp differences in the grammatical devices employed ($X^2(4, N = 650) = 36.463; p < .001$):

Table 3. Grammatical strategies in [-alt. langs]

| Strategy | Availability<br>not available | <br>available | Total |
|---|---|---|---|
| meaning shifts | 92 | 38 | 130 |
| morphosyntactic effects | 122 | 8 | 130 |
| special prosody | 106 | 24 | 130 |
| segmental effects | 120 | 10 | 130 |
| special lexicon | 99 | 31 | 130 |
| TOTAL | 539 | 111 | 650 |

## 3. Elements for a cultural evolution of ritual languages

From this, a hierarchy of ritual languages can be established as a tension between the search for Strangeness (the more the language employed departs from the normal human 'norm', the more plausible the magical powers of the user look to the community) and Learnability (the more systematic and predictable a language is, the easier its replication by a next generation of speakers):

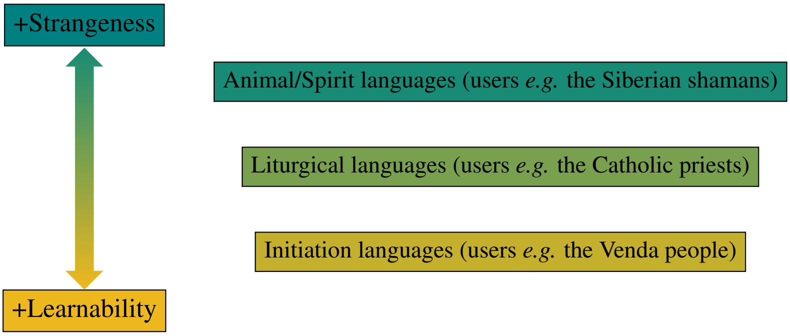


Figure 4. Hierarchy of ritual languages.

On one extreme of the hierarchy we have the shamanic practices, archetypically represented by the Siberian shamans and their idiosyncratic speech which constitutes a 'theatre of strangeness' (Eliade, 1951; Kürti, 1994; Singh, 2018). They do not show clear patterns of grammaticalization, but are full of extreme renderings with high pitch, falsettos, ventriloquism, animal sounds, cries, mumblings, etc., and the particularities of each linguistic practice typically die with the shaman. Such practices are not transmitted to the next generation, and therefore are not subject to cultural evolution pressures and dynamics (see, *e.g.* Tamariz & Kirby, 2016). At the other extreme we have community-wide practices which alter the local language with special lexicons which rather than idiosyncratic, are derived *via* transformation rules (*e.g.* the semantic and phonological antonymic patterns in Tenda ritual languages (Ferry, 1981)). As a matter of fact, ritual languages show again that 'larger communities create more systematic languages' (Raviv et al., 2019). Last, in a middle point would be the liturgical languages employed by designated individuals (and collectives) of established religious societies (priests, etc.). These are typically languages like Latin, Old Church Slavonic, Sanskrit, Pali, or Geʻez (Bennett, 2017) which require explicit teaching, and where adherence to the (archaic) norm is taken as an index of efficacy in the ritual.

Then, the grammatical devices employed for the generation of ritual variants of languages show a restricted variation: regarding meaning, they recur to the exploitation of the arbitrariness of the sign (often, with metaphors or antonymic patterns), but no new semantic composition rules are invented. Special lexicons are also employed, but they are generated by extending the meaning strategy; an exploitation of the arbitrariness of the sign. As as consequence, obscuring the message with new morpho-phonological labels for usual concepts is a very common strategy that blatantly gives the user the image of an ability that others do not possess, for he/she understands the meaning encoded by the new lexical item. The prosodic means involve generally an exploitation of what Gussenhoven (2004) terms the 'three biological codes' (the frequency code, the effort code, and the production code), but no typological shift is attested (*i.e*, no turning a stress-accent language into a tonal language for ritual purposes). Last, segmental phonological effects also seem to be mostly restricted to the inventory (the introduction of new segments, or segment substitutions), but no new vowel harmony operations. Last, the few morpho-syntactic strategies observed are circumscribed to externalization simplifications (such as the unavailability of functional elements in the ritual language, or a more productive *pro*-drop). Again, no new agreement operations or new pied-piping domains.

All in all, when humans seek a non-human pattern in their language, the 'strangeness' or 'out-of-this-world-ness' seems to be circumscribed to the lexicon and to morpho-phonology, and the patterns attested are within the limits predicted by theoretical work on natural language design (*e.g.* there are no syntactic rules that make reference to the $n^{th}$ word (Smith & Tsimpli, 1995; Musso et al., 2003), no invention of nonconservative quantifiers (Pietroski, 2005), and no syntactic rules that take as input phonological or semantic value-assignment processes (Irurtzun, 2009)). This bears testimony to the robustness of the syntactic component of human language, which in turn suggests that a large part of the variability observed across languages may be due to externalization factors (Berwick & Chomsky, 2011; Boeckx, 2014).

Even if ritual languages are intended as 'supernatural', they are canalized by natural language constraints and the biases of speakers and their communities.

## Acknowledgements

This work has received financial support from the CNRS-MITI 80|PRIME grants - 2021 PALEOSIGNES, and from the ANR ANR-21-CE27-0005 - MIND2WALL.

## References

Andersen, H. (1973). Abductive and deductive change. *Language*, *49*, 765–793.

Bennett, B. (2017). *Sacred Languages of the World.* Hoboken: Wiley.

Berwick, R. C., & Chomsky, N. (2011). The biolinguistic program: The current state of its development. In C. Boeckx & A. di Sciullo (Eds.), *The Biolinguistic Enterprise* (pp. 19–41). Oxford: Oxford University Press.

Biberauer, T., & Roberts, I. (2017). Parameter Setting. In A. Ledgeway &

I. Roberts (Eds.), *The Cambridge Handbook of Historical Syntax* (pp. 134–162). Cambridge: Cambridge University Press.

Boeckx, C. (2014). *Elementary Syntactic Structures: Prospects of a Feature-free Syntax.* Cambridge: Cambridge University Press.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329.

Eliade, M. (1951). *Le chamanisme et les techniques archaïques de l'extase.* Paris: Payot.

Ferry, M.-P. (1981). Les ganles tecreses des des ndéta (les langues secrètes des tenda). *Objets et mondes*, *21*, 173–176.

Gabb, W. M. (1875). On the indian tribes and languages of Costa Rica. *Proceedings of the American Philosophical Society*, *14*(95), 483–602.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation.* Cambridge & New York: Cambridge University Press.

Hall, K. (2019). *Shamanic Language and Verbal Art: A Comparative Examination of Hup Incantations.* Honors thesis, University of Texas, Austin.

Irurtzun, A. (2009). Why Y: On the centrality of syntax in the architecture of grammar. *Catalan Journal of Linguistics*, *8*, 141–160.

Jorgensen, E. (2020). *A Phonological Analysis of Sign Used in a Western Desert Community.* Honours thesis, University of Melbourne, Melbourne.

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review*, *24*(1), 118–137.

Kürti, L. (1994). Language, symbol and dance: An analysis of historicity in movement and meaning. *Shaman*, *1/2*, 23–82.

Lightfoot, D. (1979). *Principles of Diachronic Syntax.* Cambridge: Cambridge University Press.

Martin, J. (1817). *An Account of the Natives of the Tonga Islands in the South Pacific Ocean* (Vol. II). London: John Murray.

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature Neuroscience*, *6*(7), 774–781.

Pietroski, P. M. (2005). *Events and Semantic Architecture.* Oxford: Oxford University Press.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*, 20191262.

Singh, M. (2018). The cultural evolution of shamanism. *Behavioral and Brain Sciences*, *41*, e66.

Smith, N. V., & Tsimpli, I. M. (1995). *The Mind of a Savant.* Oxford: Blackwell.

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43.

Yang, C. D. (2002). *Knowledge and Learning in Natural Language.* New York: Oxford University Press.

# ANALYZING PROPERTY PREDICATION FROM THE PERSPECTIVE OF EVOLUTIONARY LINGUISTICS

MASAYUKI ISHIZUKA*[1] and RYOKO UNO[2]

*Corresponding Author: ishizuka@l.u-tokyo.ac.jp
[1]Department of Linguistics, University of Tokyo, Tokyo, Japan
[2]Division of Language and Culture Studies, Tokyo University of Agriculture and Technology, Tokyo, Japan

In the context of evolutionary linguistics, both constructive approaches (e.g., artificial life, robotics, and experimental semiotics) and cognitive linguistics inquire about the embodied nature of language. However, there is still a huge gap between these two areas when it comes to what constitutes language (Di Paolo et al., 2018). To bridge this gap, this paper focuses on how property predication is expressed differently among languages in a cognitive linguistic framework and argues in favor of noticing the aspect of pointing that does *not* directly contribute to establishing joint attention.

Joint attention is a proto-linguistic behavior that uses pointing with a gaze or finger to attract other people's attention to an object. Only humans can establish joint attention to share intentionality. Thus, this act is often studied as key to revealing the relationship between action and language (Gómez et al., 1993; Tomasello, 2008; 2019). Several attempts to simulate the co-evolution of joint attention and language have been made (e.g., Kwisthout et al., 2008; Uno et al., 2011; Gong & Shuai, 2012). In cognitive linguistics, demonstratives are well studied as a linguistic means to establish joint attention (Diessel, 2006). To extend the argument, this paper sees predication sentences as the linguistic equivalent of pointing. For example, if you are surprised by the redness of a flower, you might point at the flower or say, "The flower is red!"—which is a property predication sentence.

In a typological study, Stassen (1997) showed that property predication sentences have no prototypical encoding strategy of their own and they always align with one of three predicate categories: class, event, and location. Stassen characterized the prototypical strategy for each predicate category as follows: (1) The nominal strategy is prototypical for class predicates and uses a support item that is (at least etymologically) non-verbal. This strategy can be found in Waskia, for example. (2) The verbal strategy is prototypical for event predicates, uses no support item (auxiliary), shows person agreement, and has a specific negation strategy. This is exemplified by Mandarin Chinese. (3) The locational strategy is prototypical for locational predicates and uses a support item that is a verb. This strategy can be found in Finnish.

In a cognitive linguistic framework, we can interpret that property predicates have to be construed in verbal, locational, or nominal ways. Construals are how we conceptualize particular semantic content. One of the central claims in cognitive semantics is that we can portray the same situation in alternate ways (Langacker, 2008). The three strategies can be interpreted as three types of construals for the same situation.

(1) a. Nominal construal: "The flower is classified as a red thing."
    b. Verbal construal: "The flower is going through an event of redness."
    c. Locational construal: "The flower is in a location (or a state) of redness."

From (1), we make the following argument: While verbal and locational construals describe the situation in the world, nominal construal does not. Instead, nominal construal describes the inner state of the speaker. Because of this difference, what is conceptualized by sentences with the nominal construal in their immediate scope cannot be shared with others, unlike something observed in the speech context at hand. This makes nominal construals more detached from the ground than the other two types. This point is supported by another observation: many languages permit more than one encoding strategy for property predicates, and Stassen (1997, p. 615) pointed out that if a strategy switch involving the nominal strategy is interpretable in terms of time stability, the nominal strategy will encode the more time-stable option. Time stability tends to be linked to a structural description rather than to a phenomenal description (Goldsmith & Woisetschlaeger, 1982; Langacker, 1997). To elaborate, the knowledge related to the nominal strategy can be more context independent or, in fact, less grounded.

To contribute to the theory of embodied language, we interpret the analysis of property predication as an analysis of pointing in general. As in the case of property predication, pointing can be semantically ambiguous: the agent can point at something outside that can be shared to establish joint attention, but the agent can concurrently point at it to express something that is inside the agent that cannot be seen and shared with others through joint attention. The former is fully communicational, but the latter is relatively monologic. Clark (1997) points out that the communicational aspect of language attracts too much attention, often causing other aspects of language to be overlooked. An analysis of pointing in evolutionary linguistics can be regarded as an example. Our study implies that the monologic aspects of pointing must be focused on more, as they are clues to linking action to linguistic constructions with abstract meanings reflecting our structural knowledge of the world.

## References

Clark, A. (1997). *Being there: Putting brain, body and mind together*. Cambridge: MIT Press.

Di Paolo, E. A, Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The Continuity between life and language.* Cambridge: MIT Press.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics, 17,* 463–489.

Goldsmith, J. & E. Woisetschlaeger. (1982). The logic of the English progressive. *Linguistic Inquiry, 13*(1), 79–89.

Gómez, J. C., Sarria, E., & Tamarit J. (1993). A comparative approach to early theories of mind. In Baron-Cohen, S. Tager-Flusberg, H., & Cohen, D. J. (Eds.), *Understanding other minds* (pp. 195–207). Oxford: Oxford University Press.

Gong, T. & Shuai, L. (2012). Modelling the coevolution of joint attention and language. *Proceedings of the Royal Society B: Biological Sciences*, *279* (1747), 4642–4651.

Kwisthout, J., Vogt, P., Haselager, P., & Dijkstra, T. (2008). Joint attention and language evolution. *Connection Science, 20,* 155–171.

Langacker, R. W. (1997). Generics and habituals. In Athanasiadou, A. & Dirven, R. (Eds.), *On conditionals again* (pp. 191–222). Amsterdam: John Benjamins.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Stassen, L. (1997). *Intransitive predication*. Oxford: Oxford University Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny.* Cambridge: Harvard University Press.

Uno, R., Marocco, D., Nolfi, S., & Ikegami, T. (2011). Emergence of protosentences in artificial communicating system. *IEEE Transactions on Autonomous Mental Development*, *3*(2), 146–153.

# NOMINAL PLURALITY IN SINO-TIBETAN: A DIACHRONIC ACCOUNT

Jessica K Ivani*[1], Taras Zakharko[1], and Chundra Cathcart[1]

*Corresponding Author: jessica.ivani@uzh.ch
[1]Department of Comparative Language Science, University of Zürich, Zürich, Switzerland

Grammaticalization, the process by which lexical elements lose their content status to serve a grammatical function, is extensively covered in the literature on language change (Heine & Kuteva, 2002; Hopper & Traugott, 2003, among others). Several cross-linguistic qualitative studies describe this process as a unidirectional (Haspelmath, 2004; Börjars & Vincent, 2011, inter alia) and irreversible (Bybee, 2011) cline, characterized in diachronic terms by gradual and incremental steps. While existing research often conceives of grammaticalization as a whole, single process, other studies argue that it might consist of the interplay of composite and distinct processes. A body of research has focused on defining the individual variables that play a role in grammaticalization (Hopper et al., 1991; Lehmann, 2015; Petré & Velde, 2018), and recent studies have explored them quantitatively through the exploitation of corpus data (Saavedra, 2021).

In this study, we explore some aspects of the development of nominal plurality in a sample of 56 Sino-Tibetan languages, focusing on phonological and morphosyntactic integration behavior. Nominal plurality is a widespread category that has been studied in the context of grammaticalization (Heine & Kuteva, 2002). The Sino-Tibetan language family is characterized by a rich history, high variation in morphological types, and language-specific constraints that have proven to shape the course of grammaticalization processes (Bisang, 2011). Our key contribution is a better understanding of the diachronic trajectories of different variables frequently invoked in the grammaticalization literature.

The parameters used in this study describe phonological and morphosyntactic degree of cohesion of the plural markers to the respective stem(s). These variables, extracted from descriptive grammars and typological literature, include: the presence of phonological processes (PHONOLOGICAL RULES); inflectional properties such as exponence (MULTIPLE EXPONENCE), morpheme autonomy (CONCATENATIVE), and position (FIRST SLOT, FIXED POSITION). Morphosyntactic variables include adjacency (ADJACENCY) and interruptability (NOT INTERRUPTABLE). It is important to note that our analyses solely involve abstract features pertaining to integration of plural marking constructions, which can be gained

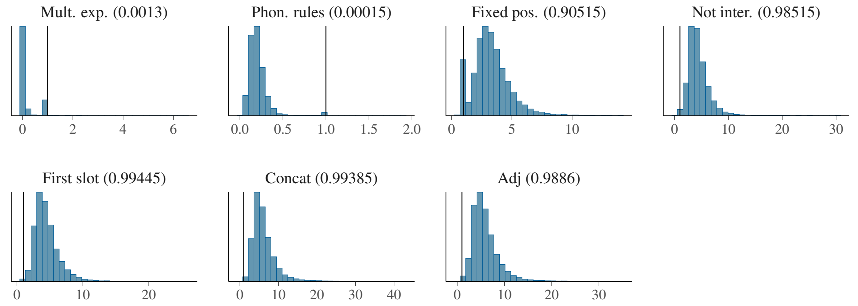


Figure 1. Posterior gain/loss ratios of different integration types, along with the proportion of samples where the ratio is greater than 1.

or lost, while grammaticalization deals with irreversible processes which embed formal elements more deeply in a language's grammar.

We use phylogenetic methods to explore whether there is a preference over time for each integration parameter (code can be found at `https://github.com/jkivani/JCoLE2022-numevo`). We model the diachronic change of each feature over the Sino-Tibetan phylogeny (Sagart et al., 2019) independently according to a continuous-time Markov process for binary data, under which a feature evolves according to a gain and loss rate. We infer posterior gain and loss rates for each feature, employing a point mass prior, under which gain and loss rates can be equal or different. We place Gamma$(1, 1)$ priors on all rates and a Beta$(1, 1)$ prior on $\pi$, the probability that gain and loss rates differ. For each posterior sample in each feature, we compute the ratio of the gain rate to the loss rate, which indicates the degree of preference for a given integration feature. These distributions are found in Figure 1, along with the proportion of samples where the ratio is greater than 1. There is decisive preference ($> .95$) for the features NOT INTERRUPTABLE, FIRST SLOT, CONCATENATIVE and ADJACENT and strong preference ($> .9$) for FIXED POSITION, and decisive dispreference ($< .05$) for the features MULTIPLE EXPONENCE and PHONOLOGICAL RULES.

In general, we see that features involving the linear order of elements are highly preferred, while phonological rules and multiple exponence are not. A possible theory is that as language families expand and absorb speakers from other families, certain features more difficult than others for second-language (L2) speakers to acquire. However, case studies suggest that phonological rules do not pose a problem for second-language speakers (Widmer, Jenny, Behr, & Bickel, 2021). At the same time, this finding does not preclude the possibility that language contact could militate against phonological and morphological complexity on a large scale. Further research from other families taking into account demographic factors is needed to obtain a richer understanding of this phenomenon.

## References

Bisang, W. (2011). Grammaticalization and linguistic typology. In *The Oxford Handbook of Grammaticalization.* Cambridge University Press.

Börjars, K., & Vincent, N. (2011). Grammaticalization and directionality. In *The Oxford Handbook of Grammaticalization.* Cambridge University Press.

Bybee, J. L. (2011). Usage-based theory and grammaticalization. In *The Oxford Handbook of Grammaticalization.* Cambridge University Press.

Haspelmath, M. (2004). On directionality in language change with particular reference to grammaticalization. *Typological Studies in Language*, *59*, 17–44.

Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization.* Cambridge University Press.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization.* Cambridge University Press.

Hopper, P. J., et al.. (1991). On some principles of grammaticization. *Approaches to grammaticalization*, *1*, 17–35.

Lehmann, C. (2015). *Thoughts on grammaticalization.* Language Science Press.

Petré, P., & Velde, F. Van de. (2018). The real-time dynamics of the individual and the community in grammaticalization. *Language*, *94*(4), 867–901.

Saavedra, D. C. (2021). *Measurements of grammaticalization: Developing a quantitative index for the study of grammatical change.* De Gruyter Mouton.

Sagart, L., Jacques, G., Lai, Y., Ryder, R. J., Thouzeau, V., Greenhill, S. J., & List, J.-M. (2019). Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences*, *116*(21), 10317–10322.

Widmer, M., Jenny, M., Behr, W., & Bickel, B. (2021). Morphological structure can escape reduction effects from mass admixture of second language speakers: evidence from Sino-Tibetan. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, *45*(4), 707–752.

# LETTERS AND THEIR SOUNDS ARE NOT PERFECTLY ARBITRARY: EXPLORING GRAPHO-PHONEMIC SYSTEMATICITY IN MULTIPLE ORTHOGRAPHY SYSTEMS

HANA JEE [1], FIONA DU [1], MONICA TAMARIZ [2], and RICHARD SHILLCOCK [1]

[*]Corresponding Author: h.jee@yorksj.ac.uk
[1]Psychology, The University of Edinburgh, Edinburgh, UK
[2]Psychology, Heriot-Watt University, Edinburgh, UK

Language, as a complex system, suggests coordination between subsystems. Recent studies demonstrated that semantically similar words tend to have similar pronunciation (Blasi et al., 2016; Dautrich et al., 2017; Jee, Tamariz, & Shillcock, 2022; Monaghan et al., 2014; Tamariz, 2008). The current research, for the first time, quantified mapping between letters and their canonical pronunciations, or *grapho-phonemic systematicity*.

We examined naturally developed phonograms (Arabic, English, Greek, and Hebrew), consciously designed phonograms (Korean, Shavian alphabet, and Pitman's shorthand), a logographic orthography (Chinese) and fictitious orthography systems (Aurebesh and Klingon).

We measured all the pairwise phonological distances between phonemes in the respective alphabet system, and the corresponding pairwise orthographical distances between letters. We then tested Pearson's *r* between these two lists of pairwise distances. The positive correlation coefficient means that similar letter-shapes have similar canonical pronunciation. In contrast, the negative correlation means that similar letter-shapes have more distinct sounds, or vice versa. We verified the significance of the correlations by conducting Monte-Carlo permutation tests.

For the phonological distance, phonemes were encoded into vectors according to the articulatory features and the distance between the vectors were calculated in various ways. We applied three methods to measure the pairwise distances between letter-shapes. *Pixel count* simply defines the distances as the difference in the number of pixels between two characters. *Perimetric complexity* is defined as ink area divided by perimeter of the character, thus the distance means the difference in complexity. *Hausdorff distance* (Huttenlocher et al., 1993) quantifies the difference between two images. Since each letter was saved as an image file (PNG), we were able to compare the contribution of the font to the grapho-phonemic systematicity.

We found the significant grapho-phonemic systematicity for all conventional writing systems and two English shorthand systems. Those fictitious alphabets did not show any systematicity. Considering each orthographic distance measure focuses on distinct aspect of the letter-shapes, the fact that a certain method maximised the systematicity of the writing system implies how it evolved.

Semitic orthography systems (Arabic, English, Greek and Hebrew) showed highest grapho-phonemic systematicity when measured by pixel count (e.g. English upper-cases $r = .22$, $p < .001$; English lower-cases $r = .14$, $p = .02$), which indicates that more articulatorily complicated phonemes take up more space in written forms. Effort in writing is easily understood as a letter's elaborateness—how long it takes to reproduce a character. Elaborateness is typically proportional to the number of pixels. Korean, Shavian alphabet and Pitman's shorthand were all intentionally designed to exploit the systematicity between letters and sounds. For instance, voiced-voiceless phoneme pairs share the identical visual features with slight variations. This topological difference was well-captured by Hausdorff distance (e.g. Korean KCC Eun-young $r = .39$, $p <. 001$).

Although limited in number (N = 58), we found the significant grapho-phonemic systematicity in the Chinese characters that are acquired in the first and second year of the primary school. We found the negative correlation coefficient ($r = - .12$, $p < .001$), indicating that Chinese was influenced by an evolutional force that *distinguishes* linguistic symbols. The finding implies that grapho-phonemic systematicity may exist to facilitate language learning and orthography acquisition.

Our analyses are first a proof of concept: it is possible to quantify grapho-phonemic systematicity across a whole alphabet, for particular fonts and for different languages. We also have confirmed and quantified the systematicity intended by the authors of Korean writing system and English shorthand systems. Our future research can shed more lights on sub-structure of grapho-phonemic systematicity: the contribution of each phoneme/letter to the whole systematicity; whether the more frequent phoneme/letter contributes more to the whole systematicity; and most importantly, how this grapho-phonemic systematicity bootstraps infants' learning orthography.

## References

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818-10823.

Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive science*, *41*(8), 2149-2169.

Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, *15*(9), 850-863.

Jee, H., Tamariz, M., & Shillcock, R. (2022). Exploring meaning-sound systematicity in Korean. *Journal of East Asian Linguistics, 31*(1),

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130299.

Tamariz, M. (2008). Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *The Mental Lexicon*, *3*(2), 259-278.

# PHYLOGENETIC MULTILEVEL MODELS REVEAL A SIMPLICITY BIAS IN URALIC

Yingqi Jing*[1], Miina Norvik[23], Outi Vesakoski[4], and Michael Dunn[1]

*Corresponding Author: yingqi.jing@lingfil.uu.se
[1]Department for Linguistics and Philology, Uppsala University, Uppsala, Sweden
[2]Department of Modern Languages, Uppsala University, Uppsala, Sweden
[3]Institute of Estonian and General Linguistics, University of Tartu, Tartu, Estonia
[4]Department of Finnish language and Finno-Ugric linguistics, University of Turku, Turku, Finland

One central topic in comparative linguistics is to investigate the evolutionary rates of typological traits between different levels of language systems. Varying rate of change could take place due to functional constraints such as general principles of efficiency in communication and learning (Zipf, 1949; Culbertson & Kirby, 2016; Hahn, Jurafsky, & Futrell, 2020), or language contact and second language acquisition that may affect the linguistic complexity (McWhorter, 2011; Lupyan & Dale, 2010; Bentz & Berdicevskis, 2016; Housen, De Clercq, Kuiken, & Vedder, 2019). Previous research on rate variation across domains is however unclear. Greenhill et al. (2017) and Carling and Cathcart (2021) show that grammatical features tend to evolve faster than lexical items (i.e., basic vocabulary) and morphological features at least in Austronesian and Indo-European, whereas Dediu and Cysouw (2013) show that syntactic features including word orders are more stable than morphological features, or there is no much rate difference across domains (Greenhill, Atkinson, Meade, & Gray, 2010). Newly available Uralic typological data allows us to estimate the rates of change across typological domains (Norvik et al., 2022) and understand the evolutionary dynamics of the Uralic language system.

With the publications of the Uralic Typological Database (Norvik et al., 2022), we use Bayesian phylogenetic inference to estimate the rates of change for different typological features in the history of Uralic. We go beyond the earlier approaches to fit phylogenetic models for each individual feature or estimate the average rates across all features. Instead, we introduce a novel multilevel phylogenetic Continuous-time Markov Chain model to investigate the evolutionary trajectories of 110 features across different domains (phonology, morphology and syntax) in 33 Uralic languages. The hierarchical model allows us to jointly infer the evolutionary rates at both population and group levels, guarding against overfitting and underfitting (Nalborczyk, Batailler, Lœvenbruck, Vilain, & Bürkner, 2019; Stan Development Team, 2021). We first validate the model with simulated data and then apply our approach to Uralic typological data.

Our results reveal a slight directional change towards simple character states at the population level (mean rate of $q_{01} = 6.8$ and 95% CI = [0.92, 23]; mean rate of $q_{10} = 8.26$ and 95% CI = [0.76, 24.1]), suggesting that losing a complex feature on average takes around 250 years less time than gaining a complex one. In each domain, we also observe consistent simplicity biases, though the estimated rates of change are quite similar across different domains (mean rate in phonology: 8.1 and 95% CI = [0.79, 30.3]; mean rate in morphology: 7.29 and 95% CI = [0.52, 27.9]; mean rate in syntax: 8.6 and 95% CI = [0.47, 35]).

The evolutionary biases towards simplified states can be driven by the general economic principle, which reduces the linguistic complexity to facilitate communication (Zipf, 1949). The observed trends are quite consistent across domains, suggesting that forces of simplification are persistent in the whole Uralic language system. The simplification of language systems would also be expected in certain contact situations (e.g., imperfect learning or language shift), which may lead to the loss of complexity or redundancy (Heine & Kuteva, 2005; Grünthal et al., 2022). It is also worth noting that even though the differences in rate biases are slight, these evolutionary preferences can be amplified in a long period of language evolution or learning. Our results are not consistent with previous work that suggests unequal rates of change across domains in other languages families. Instead, we show that language evolves at very similar rates across typological domains in Uralic. Further research is needed to expand our approach to other families to see whether there is a constant rate constraint in the multilayer system of languages (Kroch, 1989; Kauhanen & Walkden, 2018).

## References

Bentz, C., & Berdicevskis, A. (2016). Learning pressures reduce morphological complexity:: Linking corpus, computational and experimental evidence. In *26th international conference on computational linguistics* (pp. 222–232). Osaka, Japan.

Carling, G., & Cathcart, C. (2021). Reconstructing the Evolution of Indo-European Grammar. *Language*, *97*(3), 561–598.

Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, *6*(1964).

Dediu, D., & Cysouw, M. (2013). Some Structural Aspects of Language Are More Stable than Others: A Comparison of Seven Methods. *PLoS ONE*, *8*(1), e55009.

Greenhill, S. J., Atkinson, Q. D., Meade, A., & Gray, R. D. (2010). The shape and tempo of language evolution. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1693), 2443–2450.

Greenhill, S. J., Wu, C.-H., Hua, X., Dunn, M., Levinson, S. C., & Gray, R. D.

(2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*, 201700388.

Grünthal, R., Heyd, V., Holopainen, S., Janhunen, J., Khanina, O., Miestamo, M., Nichols, J., Saarikivi, J., & Sinnemäki, K. (2022). Drastic demographic events triggered the uralic spread. *Diachronica (to appear).*

Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, 201910923.

Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change.* Cambridge University Press.

Housen, A., De Clercq, B., Kuiken, F., & Vedder, I. (2019). Multiple approaches to complexity in second language research. *Second language research*, *35*(1), 3–21.

Kauhanen, H., & Walkden, G. (2018). Deriving the Constant Rate Effect. *Natural Language & Linguistic Theory*, *36*(2), 483–521.

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, *1*(3), 199–244.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1), e8559.

McWhorter, J. H. (2011). *Linguistic simplicity and complexity: Why do languages undress?* Walter de Gruyter.

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An introduction to bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242.

Norvik, M., Jing, Y., Dunn, M., Forkel, R., Honkola, T., Klumpp, G., Kowalik, R., Metslang, H., Pajusalu, K., Piha, M., Saar, E., Saarinen, S., & Vesakoski, O. (2022). Uralic typology in the light of new comprehensive data sets. *Journal of Uralic Linguistics*, *1*(1), 4-42.

Stan Development Team. (2021). *RStan: the R interface to Stan.* (R package version 2.21.2)

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort.* Cambridge, MA: Addison-Wesley Press.

# POINTER EVOLUTION POINTS TO GRADUAL EVOLUTION OF HIERARCHICAL COMPLEXITY

SVERKER JOHANSSON [*1]

[*]Corresponding Author: sja@du.se
[1]Dalarna University, Falun, Sweden

Chomsky (e.g. 2010) and others regard unlimited Merge as the defining feature of language, that cannot evolve gradually. The neural implementation of Merge is not well understood (Rizzi 2012, Zaccarella et al 2017), but must involve something functionally equivalent to pointers in working memory. Every Merge requires two pointers, and full syntactic trees may require dozens. Other syntactic paradigms also need pointers.

Humans do hierarchies in general better than chimpanzees. Any hierarchical thinking requires nested pointers in working memory, but they are neurologically expensive and degrade with depth (Crawford et al. 2016). Humans have larger working-memory capacity than chimpanzees, which has been proposed as key to human cognitive evolution (Read 2008, Coolidge & Wynn, 2005). Gradual evolutionary growth of pointer capacity will allow gradually increasing syntactic complexity, without saltations in the underlying computational machinery. Both depth degradation and pointer capacity naturally limit Merge even in modern humans, consistent with corpus data (e.g. Karlsson 2010).

## 1. Can infinity evolve?

Language is commonly said to be infinite, and this is true at least in the limited sense that there are no limits to what can be said. But some popular linguistic paradigms, notably minimalism (Chomsky 1995), postulate that language is inifinite in a stronger, more literal sense, in that the language faculty can, in principle, generate an infinite number of infinitely long sentences. The generating hierarchy-building operation (Merge, in Chomsky's case) can build trees of arbitrary depth, without limits. This kind of inifinity is sometimes invoked as an argument against the gradual evolution of the human language capacity, roughly along the following lines (e.g. Chomsky 2010, Berwick & Chomsky 2015):

1. Language is based on Merge, which is the defining feature of language.
2. Merge is unlimited.
3. Language, generated by unlimited Merge, is infinite.

4. There is no gradual path from the finite to the infinite, no possible intermediate “semi-infinite” proto-language.
    a. Having a limited Merge, and then gradually evolving an increased limit, will not get us to infinity.
5. Merge, and thus language, must have arisen in a single step. There is no such thing as “half-Merge”, no intermediate stage between no Merge and full Merge (Berwick & Chomsky 2019).
6. *Conclusion*: Language did not evolve gradually.

But humans are indubitably descended from language-less ape-like ancestors. And the notion that something as complex as language could have arisen in one fell swoop, by a single super-mutation, is not tenable (e.g. Tallerman 2014, de Boer et al. 2019).

If the human language capacity did evolve in several steps, one of the points above must be incorrect. The question is which one.

## 2 Merge is limited

I will focus first here on assumption #2, that Merge is unlimited, and show that this is not a correct description of the actual human language capacity.

The term “Merge” *sensu stricto* is specific to Chomskyan minimalism, but regardless of linguistic paradigm there will be a need for some kind of hierarchy-handling neural machinery in the language capacity, as human languages indubitably do have some hierarchical structure. The form that the hierarchies take differ widely between paradigms, but my argument here is intended to apply over a broad range of hierarchy-handlers. I will use “Merge” here as a label, both because it is the best-known hierarchy handler, and because most proponents of the infinity argument above are Chomskyans, but I use it *sensu lato*, as a label for a generic hierarchy-handler.

From an empirical perspective, language is clearly finite. The human brain has a finite size, and the human lifespan is finite, so infinite production is obviously impossible in practice. Actual language usage shows that Merge in practice is not just limited, but limited to fairly shallow depths – in written corpus data, it is exceedingly rare to find examples of phrasal embedding more than three layers deep, and the limits appear even tighter in spoken corpora (Karlsson 2010). Similarly, the accuracy of grammaticality judgements approaches chance level as embedding depth increases beyond what is commonly used (Christiansen & MacDonald 1999).

These performance limits keep real human languages from being infinite. Languages are still very large; a few levels of embedding combined with a normal vocabulary still allows for an astronomical number of different sentences – unlimited for all practical purposes but not infinite.

However, it has been argued since Chomsky (1965) that performance is not interesting, that performance limitations just distract attention from the real underlying linguistic competence. The latter is what linguists should study, and the latter is postulated to be infinite.

But is competence, a theoretical entity that is never directly observed, actually the proper target for language *evolution* studies? Isn't it enough to account for the evolution of actual language usage?

## 3 How can Merge be implemented in the brain?

The neural implementation of Merge is not well understood (Rizzi 2012, Zaccarella et al. 2017). Discussions in the literature are mainly about the computational machinery, but I will focus instead on memory needs, that are non-trivial. In order to build a syntactic structure in the brain, two types of objects need to be stored in memory: (1) lexicon storage in long-term memory (LTM), and (2) syntactic nodes in working memory (WM). LTM and WM are distinct types of memory, with distinct characteristics (Norris 2017). The leaf nodes in the syntactic tree in WM must somehow refer to lexical items in LTM. Unless entire lexical items are copied into WM nodes, this must involve something functionally equivalent to *pointers* (e.g. Reilly 2003) in working memory, where a WM node contains a reference to a LTM item (Takac & Knott 2016); this would be 0-merge *sensu* Rizzi (2016).

A Merge operation will create a new object in WM, which consists of two pointers, one to each object that is merged (plus features and whatever else is stored at each node). If both pointers refer to lexical items in LTM, we have 1-merge *sensu* Rizzi (2016). The brain must be able to handle and refer to this new composite object as a single entity, a *chunk* of memory (cf. Gobet et al. 2016, Isbilen & Christiansen 2018), for purposes of further merging; chunking is a prerequisite for Merge.

For the next level of merging, it is not enough with pointers from WM to LTM; pointers from one WM location to another WM location are also needed. WM-to-WM pointers are likely neurally distinct from WM-to-LTM pointers, as the address space is different in kind. This means three different types of Merge nodes are needed in WM:

- Merging two lexical items (1-merge of Rizzi 2016).
- Merging a merged item with a lexical item (2-merge of Rizzi 2016).
- Merging two merged items (3-merge of Rizzi 2016).

All this is done in WM, and it can be noted that, while many animals do have working memory (Carruthers 2013), there is fair evidence that humans have more of it than even our closest living relatives (Read 2008, Coolidge & Wynn, 2005).

In recent works, Chomsky does note the need for working memory (or "workspace", as he calls it), for syntactic processing, but does not discuss either implementation or limitations (e.g Chomsky et al. 2017).

## 4 Pointers

Pointers – having one memory location contain a reference to another memory location – are used extensively in computer programming. Biological brains are doing many operations for which a computer would use pointers. Something functionally equivalent to pointers must be neurally implemented.

### *4.1 Who needs pointers?*

Pointers in the brain are needed as soon as the brain in one location manipulates information that is stored elsewhere in the brain. Notably, any WM computations involving LTM items will require pointers. It is rather pointless to have both types of memory, unless you have pointers as well. But both WM and LTM are widespread among non-human animals. Fish have both WM (Hughes & Blight 1999) and LTM (Lucon-Xiccato & Bisazza 2017), so presumably all vertebrates do, though most likely not identical to human memory in capacity or capabilities (Carruthers 2013). Pointers, at least WM-to-LTM, thus have an ancient origin. This means that 0-merge (Rizzi 2016) is available to all vertebrates. For 1-merge (or higher), chunking is required, for which there is evidence in e.g. rats (Fountain & Benson 2006) as well as some other mammals and birds, but negative results for fish and amphibians (Wickelgren 1979), suggesting a more limited distribution that nevertheless includes many (all?) non-primate mammals.

For 2+merge, WM-to-WM pointers are needed. Any WM operations involving hierarchical structures would be evidence of WM-to-WM pointers. There is some evidence of hierarchical cognition in non-human primates (e.g. Seyfarth et al. 2005), but it is not strongly compelling. The jury is still out on 2+merge in non-humans.

Humans, however, are hierarchical thinkers *par excellence*, to the extent that Fitch (2014) labels us "dendrophiles" for our propensity to use hierarchical thinking and impose hierarchical structure on anything and everything. Martin & Doumas (2017) propose that this general mechanism for thinking hierarchically can be repurposed for linguistic structures.

### *4.2 Neural implementation of pointers*

Pointers are inherently difficult to handle in the brain, as memory addressing is not a matter of just storing the number of the addressed memory cell, like it is in a computer. The content addressable memory in the brain requires a fundamentally different type of pointers, that are neurally quite expensive. This is particularly true when attempts are made to scale up proposed pointer models to a human-sized address space; most models do not scale well and cannot address a human-sized memory with the number of neurons available in a human brain (Blouw et al. 2016). Exact pointers are particularly vulnerable to scaling issues, whereas different types of approximate pointers fare better (Crawford et al. 2016, Legenstein et al. 2016).

The model of Crawford et al. (2016) is attractive in this context, as it has explicitly been shown in simulation to manage the full human lexicon with a reasonable number of neurons that will actually fit within the relevant brain areas. This model is based on lossy compression of information; accuracy remains adequate for single pointers, but degrades rapidly with depth when pointers are nested in recursive structures; the degradation mimics actual human performance (as opposed to theoretical competence) on multi-level embeddings.

But pointers remain expensive and consume WM fast, especially if you have hierarchical structures with multiple pointers-to-pointers.

## 5 The gradual evolution of limited Merge

As reviewed in the previous section, many non-human animals have LTM and WM, of limited size, as well as WM-to-LTM pointers, which implies 0-merge. Chunking, and thus 1-merge, likewise can be found in a fair range of animals (Wickelgren 1979), though the evidence for 2+merge outside the human lineage is more limited.

Contrasting this with the capacities of modern humans, we can conclude that the evolutionary changes in the human lineage, after we parted ways with the other apes, most likely include:

- Expanded WM.
- Expanded LTM (including possibly dedicated lexical storage).
- Dendrophilia.
- WM-to-WM pointers, and likely generally enhanced pointer handling.
- Node structures for 2-merge and 3-merge.
- Nested pointer handling.
- Enhanced chunking?

None of these changes need to be linguistically motivated; they are plausible components in the general human cognitive enhancement that took place concurrently (cf. Sherwood et al. 2008). There has obviously been considerable selective pressure along the human lineage for enhanced brain size and presumable enhanced cognition (Bailey & Geary 2009). Such a selective pressure explains the evolution of expanded working memory, as there is a strong correlation between WM size and general intelligence (e.g. Colom et al. 2008).

Given the WM cost of hierarchical structures, the WM expansion will make a big difference in hierarchy handling capacities. With a small WM, you cannot build any significant tree structures even if full-blown Merge is computationally available. A larger WM invites the evolution of enhanced pointer handling, including WM-to-WM pointers and recursively nested pointers, which were pointless before. There is also more scope for chunking in WM, when there is room for more than a few chunks. Likewise, with a larger WM, dendrophilia (Fitch 2014) starts making sense.

This leaves us with a proto-human who has a fair-sized WM, a basic set of pointer operations, including WM-to-WM pointers, and a general Merge-like operation that can do all the merge levels of Rizzi (2016). Such a proto-human can combine lexical items into two-word phrases, and can combine two phrases into a composite utterance. Dedicated syntactic machinery such as feature-checking is still missing, leading to a rather anarchic proto-language, but the basic hierarchical structure is there. Pointers remain limited in both address space and nesting depth, imposing limits on both lexicon size and tree size. But both limits can be relaxed simply by gradually adding more neurons to the pointer machinery.

If there is selective pressure towards more expressive language, with more complex syntax and enhanced narrative capacities, this can be dealt with in at least three ways:

- Further WM expansion.
- Pointer expansion as above, to handle both a human-sized address space and the corpus-attested (Karlsson 2010) nesting depth.
- The general Merge operation already available may be augmented with language-specific add-ons, gradually adding all the operations that modern Merge does beyond the actual merging (feature-checking etc.), in order to make linguistic processing more precise and efficient.

The end result is modern humans, with a modern human language capacity. Merge, or whatever hierarchical operation is actually running the human language capacity, has gradually been enhanced to its modern full-fledged form, with no infinity paradoxes blocking the way.

## References

Berwick, Robert C. & Chomsky, Noam. (2015) *Why only us?* Cambridge, MA: MIT Press.

Berwick, Robert C. & Chomsky, Noam. (2019) All or nothing: No half-Merge and the evolution of syntax. *PLOS Biology* 17:e3000539.

Bailey, D.H., Geary, D.C. (2009) Hominid Brain Evolution. *Hum Nat* **20,** 67–79.

Blouw et al. (2016) Concepts as semantic pointers: A framework and computational model. *Cognitive Science* 40:1128-1162

Carruthers, Peter (2013) Evolution of working memory. *Proc Nat Acad Sci* 110:10371-10378.

Chomsky, Noam. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press

Chomsky, Noam. (2010). Some simple evo devo theses: how true might they be for language? In Larson et al. (Eds.), *The Evolution of Human Language. Biolinguistic Perspectives*. Cambridge: Cambridge University Press.

Chomsky, Noam. (1995). *The minimalist program*. Cambridge: MIT Press

Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. (2017). *Generative grammar and the faculty of language: Insights, questions, and challenges.* Unpublished manuscript. https://ling.auf.net/lingbuzz/003507

Christiansen, Morten H. & MacDonald, Maryellen C. (1999) *Processing of recursive sentence structure: Testing predictions from a connectionist model.* http://www.academia.edu/download/39800748/Processing_of_Recursive_Sentence_Structu20151108-5695-17yb6hj.pdf (retrieved 2019-09-11)

Colom et al. (2008) Working memory and intelligence are highly related constructs, but why? *Intelligence*, 36:584-606.

Coolidge, Frederick L & Wynn, Thomas (2005) Working memory, its executive functions, and the emergence of modern thinking. *Cambridge Archaeological Journal* 15:5-26.

Crawford, Eric & Gingerich, Matthew & Eliasmith, Chris (2016) Biologically plausible, human-scale knowledge representation. *Cognitive Science* 40:782-821.

de Boer, Bart, Bill Thompson, Andrea Ravignani & Cedric Boeckx. (2019) *Evolutionary dynamics do not motivate a single-mutant thoery of human language.* bioRxiv preprint 517029 https://www.biorxiv.org/content/10.1101/517029v1

Fitch, W Tecumseh (2014) Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys Life Reviews* 11:329-364

Fountain, Stephen B. & Benson, Don M. (2006) Chunking, rule learning, and multiple item memory in rat interleaved serial pattern learning. *Learning and Motivation,* 37:95-112.

Gobet, F. & Lloyd-Kelly, M. & Lane, P. C. R. (2016) What's in a name? The multiple meanings of 'chunk' and 'chunking', *Frontiers in Psychology* 7:102

Hughes, Roger N. & Blight, Christine M. (1999) Algorithmic behaviour and spatial memory are used by two intertidal fish species to solve the radial maze. *Animal Behaviour*, 58:601-613.

Isbilen, E S & Christiansen, M H (2018) Chunk-based memory constraints on the cultural evolution of language. *Topics in Cognitive Science* 2018:1-14

Karlsson, Fred (2010) Syntactic recursion and iteration. In Harry van der Hulst, ed., *Recursion and Human Language.* Berlin/New York: Mouton de Gruyter,

Legenstein et al. (2016) Assembly pointers for variable binding in networks of spiking neurons. q-bio.NC/1611.03698 (https://arxiv.org/abs/1611.03698 )

Lucon-Xiccato, Tyrone & Bisazza, Angelo (2017) Complex maze learning by fish. *Animal Behaviour*, 125: 69-75.

Norris, Dennis (2017) Short-term memory and long-term memory are still different. *Psychological Bulletin* 9:992-1009

Martin, A. E. & Doumas, L. A. A. (2017) A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biology* 15:e2000663.

Read, Dwight W (2008) Working memory: A cognitive limit to non-human primate recursive thinking prior to hominid evolution. *Evolutionary Psychology* 6:676-714.

Reilly, Edwin D. (2003). *Milestones in Computer Science and Information Technology.* Westport: Greenwood Press.

Rizzi, Luigi (2012) Core linguistic computations: How are they expressed in the mind/brain? *Journal of Neurolinguistics* 25:489-499.

Rizzi, Luigi. (2016) Monkey morpho-syntax and merge-based systems. *Theoretical Linguistics* 42(1-2): 139–145

Seyfarth, Robert M. & Cheney, Dorothy L. & Bergman, Thore J. (2005) Primate social cognition and the origins of language. *Trends in Cognitive Sciences*, 9: 264-266.

Sherwood, C.C., Subiaul, F. and Zawidzki, T.W. (2008), A natural history of the human mind: tracing evolutionary changes in brain and cognition. *Journal of Anatomy*, 212: 426-454.

Takac, M & Knott, A (2016) *Mechanisms for storing and accessing event representations in episodic memory, and their expression in language: A neural network model.* In Proceedings of the 38th Annual Conference of the Cognitive Science Society

Tallerman, Maggie. (2014) No syntax saltation in language evolution. *Language Sciences* 46: 207-219

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review, 86*(1), 44–60.

Zaccarella et al (2017) Building by syntax: the neural basis of minimal linguistic structures. *Cerebral Cortex* 27:411-421.

# THE COLOUR LEXICON IS SHAPED BY ENVIRONMENT AND BIOLOGY: COMPARING HIMBA AND FRENCH COLOUR PERCEPTION

Mathilde JOSSERAND*[1], Serge CAPAROS[2,3], François PELLEGRINO[1], and Dan DEDIU[4]

*Corresponding Author: mathilde.josserand@univ-lyon2.fr
[1]Laboratoire Dynamique du Langage, CNRS & University Lumière Lyon 2, Lyon, France
[2] Department of Psychology, University of Paris 8, Saint-Denis, France
[3]Institut Universitaire de France, Paris, France
[4] University of Barcelona & Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain

The evolution and diversification of language is driven by evolutionary pressures, and by its repeated use, learning, and transmission across generations. One aspect of this process is the amplification of weak biases, potentially up to a whole language and beyond, in large-scale cross-linguistic patterns. For example, differences in subsistence mode (Blasi et al, 2019), in the anatomy of the vocal tract (Dediu et al, 2019), or in climate (Everett et al., 2015) may influence the phonemic repertoires of languages. We show here that the lexicons of languages can be influenced by weak biases due to environmental factors. UV radiations affect the eye physiology by causing the lens to opacify, and this may reduce the ability to perceive the blue part of the colour spectrum (Davies et al, 1998). As a consequence, people living in areas with high levels of UV incidence may be less likely to possess a specific term for blue in their vocabulary (Lindsey & Brown, 2002). However, the causes underlying the evolution of colour lexicons are complex and likely involve multiple factors, including variation in other cultural and physical factors. To investigate whether Lindsey and Brown's physiological hypothesis still holds when these other factors are considered, Josserand et al. (2021) conducted a large-scale statistical study on 142 populations, and found that populations living in areas with high levels of UV incidence are more likely to merge 'green' and 'blue' colours under a single (*'grue'*) term.

Here, we wished to complement these results by investigating the relationship between colour perception and UV incidence in two representative populations living in markedly different environmental conditions. Specifically, we compared French participants (who are exposed to a medium amount of UV radiation) to Himba participants (a population of Northern Namibia, living mostly outdoor under high UV incidence). While French speakers have a specific term for ‘blue’, the Himba historically do not (Roberson et al., 2004; but see Mylonas et al., 2022). Here, we present new preliminary results from 76 Himba and 42 French participants. We used the Farnsworth Munsell test, which evaluates and ranks colour acuity for a broad range of colours and can reveal both inherited and acquired colour vision deficiencies. In addition, we used the JND (just noticeable differences) task for blue and red colours. Here, participants performed a binary forced choice stimulus task using a computerized adaptive testing design. Two coloured rectangles were presented on the screen with a colour difference magnitude ranging from 0 to 10 CIELAB units and applied differently per colour and per axis (L, a and b). For both experiments, luminosity settings and screen calibration were carefully controlled.

The results showed that older French participants were less able to discriminate close colours in the blue range (only). This finding is consistent with the cumulative nature of the exposure to UV radiation throughout lifespan. Second, older Himba participants were less able to discriminate close colours *both* in blue and red ranges. Surprisingly, while we expected that the difference between Himba and French participants would be the strongest in the blue range, on the contrary, Himba participants had greater difficulty discriminating close colours through the whole range of colours. However, a careful analysis is in progress to account for potential confounding factors, for instance related to the inadequacy of the tasks themselves.

These preliminary results may suggest that the effect of UV radiations on colour perception occurs for all hues. If so, it may be one factor explaining the relatively low number of colour terms used by the Himba people, by limiting the hues individualized in the speakers’ perceptual space, in addition to other cultural and environmental factors. Indeed, recent computational work suggests that both perceptual structure and communicative needs shape colour naming (Zavlasky et al, 2019), and injecting variation in the initial perceptual space of these models is an avenue we plan to explore in the future. Our contrastive study between two populations contributes to unravelling the role of colour perception in mediating physical/environmental factors on lexical characteristics of languages. Overall, this work suggests that language is deeply intertwined with its surroundings, thus highlighting the relevance of considering weak biases related to the socio-cultural system, the features of the environment, or the biology of its speaker.

## Acknowledgements

We thank our two guides and translators Fanny Kavary and Chinho Lopez for their great professionalism and help with the experiments, as well as for sharing their knowledge on Himba culture. We also want to acknowledge Esther Boissin for her resourceful advice and emotional support in this fieldwork. Most importantly, we are immensely grateful to the Namibians who welcomed us and our project with openness, benevolence, and generosity.

## References

Blasi, D. E., Moran, S., Moisik, S. R., Widmer, P., Dediu, D., & Bickel, B. (2019). Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science*, 363(6432).

Davies, I. R. L., Laws, G., Corbett, G. G., & Jerrett, D. J. (1998). Cross-cultural differences in colour vision: Acquired ‘colour-blindness’ in Africa. *Personality and Individual Differences*, 25(6), 1153-1162.

Dediu, D., Janssen, R., & Moisik, S. R. (2019). Weak biases emerging from vocal tract anatomy shape the repeated transmission of vowels. *Nature Human Behaviour*, 3, 1107–1115.

Everett, C., Blasi, D. E., & Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences*, 201417413.

Josserand, M., Meeussen, E., Majid, A., & Dediu, D. (2021). Environment and culture shape both the colour lexicon and the genetics of colour perception. *Scientific Reports*, 11(1), 19095.

Lindsey, D. T., & Brown, A. M. (2002). Color naming and the phototoxic effects of sunlight on the eye. *Psychological science*, 13(6), 506–512.

Mylonas, D., Caparos, S., & Davidoff, J. (2022). Augmenting a colour lexicon. *Humanities and Social Sciences Communications*, 9(1), 1–12.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2004). The Development of Color Categories in Two Languages: A Longitudinal Study. *Journal of Experimental Psychology: General*, 133(4), 554-571.

Zaslavsky, N., Kemp, C., Tishby, N., & Regier, T. (2019). Color Naming Reflects Both Perceptual Structure and Communicative Need. *Topics in Cognitive Science*, 11(1), 207-219.

# LANGUAGE WITHOUT SHARED MEANING: PRELIMINARY CONSIDERATIONS ON THE EVOLUTIONARY BENEFITS OF SUBJECTIVITY

Alexandre Kabbach*[1,2] and Aurélie Herbelot[2,3]

*Corresponding Author: alexandre@kabbach.net
[1]Department of Linguistics, University of Geneva, Geneva, Switzerland
[2]Center for Mind/Brain Sciences, University of Trento, Trento, Italy
[3]Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

We propose a characterization of language that does not rest on the hypothesis that meaning is necessarily shared across interlocutors, since it is fundamentally grounded in the privacy and subjectivity of mental content. We first argue that the function of language is thus best characterized as the *coordination* rather than the *communication* between minds, since it is the constraints on *use*—rather than on meaning *itself*—that are negotiated between interlocutors during linguistic interactions. We then explore the evolutionary benefits of subjectivity and argue that it positively contributes to *adaptability* through: 1) *innovation*, as conceptual variability at the individual level increases the likelihood of the group of finding relevant conceptual innovations when exposed to environmental challenges; and 2) *transfer*, as subjective coordination allows for individual innovations to spread across a whole population, without individuals having to align their conceptual spaces and thereby lose the benefits of conceptual variability.

## 1. Introduction

Communication through language rests on mutual understanding: to understand you, I must figure out exactly what you *mean* by what you *say*. Successful communication therefore requires shared meaning and appeal to mental content. And yet, mental content remains inherently *private* and *idiosyncratic* in nature, which poses a major challenge to communication, for how can we indeed guarantee mutual understanding if the meaning speakers intend to communicate through language remains highly subjective and inaccessible to others (Pelletier, 2017, p.63)?

In this work, we propose to turn the problem on its head: rather than trying to reconcile subjectivity with communication, we explore what it would mean for the characterization of language to be structured around subjectivity, and what benefits subjectivity could have from an evolutionary perspective.

Our first contribution is to argue that the function of language that unfolds is actually best characterized as the *coordination* rather than the *communication* between minds—where communication and coordination are formally distinguished with respect to their requirement of shared meaning: strict for communication, loose for coordination. Concretely, we argue that coordinating minds is best conceived as the *private satisfaction of shared constraints*, given that 1) speakers never have direct access to what their interlocutors actually mean, so that; 2) it is the constraints on *use*—rather than on meaning *itself*—that are actually negotiated between interlocutors during linguistic interactions.

Our second contribution is to formalize a clear hypothesis about the evolutionary benefits of subjectivity in improving adaptability through cumulative culture—that is, the ability of humans to innovate by creating new knowledge as well as to preserve existing knowledge by passing it on to other members of the group. We argue that subjectivity is advantageous in two ways: 1) for *innovation*, conceptual variability at the individual level increases the likelihood of the group of finding relevant conceptual innovations when exposed to environmental challenges; and 2) for *transfer*, subjective coordination allows for individual innovations to spread across a whole population, without individuals having to align their conceptual spaces and thereby lose the benefits of conceptual variability.

Through this work, we hope to contribute to the discussions on both the evolution and the function of language by decoupling the characterization of *communication success* from the characterization of *interpretation success*. Doing so allows us to treat the communicative function of language no longer as a theoretical prerequisite, "hardcoded" into our model of language, but as a mere hypothesis that can now be analyzed comparatively for explanatory adequacy.

## 2. Language, communication and the problem of subjectivity

### 2.1. *The code model of communication*

According to Sperber and Wilson (1986/1995, p.2) "From Aristotle through to modern semiotics, all theories of communication were based on [. . .] the code model" later formalized by Shannon and Weaver (1949), which characterizes communication as information transfer between a sender and a recipient who exchange messages encoded and decoded to and from (linguistic) signals through a potentially noisy channel.

The problem with such a characterization—as *pragmatics* has extensively shown—is that *there is more to language than what is said*, so that linguistic signals alone often do not suffice to unambiguously decode messages: they must be put into *context* (Sperber & Wilson, 1986/1995, ch.1 §2). But context here is to be understood not only as the linguistic (e.g. sentential or discourse) context under which communication takes place, but also as the non-linguistic context such as the set of background experiences, knowledge, beliefs, desires or assumptions grounding the interpretation of messages and the mapping of linguistic *form* to conceptual *meaning*. Recipients must understand not only *what is said*, but also what senders (or speakers) actually *mean* by what they say, and to do so they must identify their *intentions* (Grice, 1989).

Inferential processes are therefore required to properly decode messages, and decoding is probably best formalized as *interpretation*, given how messages are themselves best characterized in terms of mental content and as combinations of propositions and propositional attitudes (Sperber & Wilson, 1986/1995, p.57).

### 2.2. *Subjectivity, similarity and communication success*

Major problems then arise when trying to characterize communication success. Because mental content is inherently private and background experiences grounding conceptual knowledge necessarily idiosyncratic and thus highly subjective and speaker-specific (Pelletier, 2017, §6). Yet, the code model assumes that communication is successful when the message decoded by the recipient is *identical* to the one encoded by the sender. How can we guarantee, then, that senders and recipients do in fact understand one-another if the messages they map to and from identical linguistic signals are processed through *different* conceptual backgrounds?

The problem is well acknowledged[1] and usually resolved along two lines. First, by relaxing the identicity constraint on messages to sheer similarity. Second, by emphasizing that agents in linguistic interactions are not passive but actively negotiate meaning and gradually align their conceptual representations through conversation (Clark, 1996; Pickering & Garrod, 2006). But what does it mean for two messages to be similar? And how much similarity is enough to guarantee communication success? In effect, the concept of similarity merely displaces the problem while still posing major theoretical and empirical challenges (Medin, Goldstone, & Gentner, 1993). Alignment-based approaches to communication are not without problems either, as they still remain committed to the code model's characterization of communication success, inasmuch as conceptual representations are expected to align across interlocutors during conversation and misalignment between speakers is only tolerated when it pertains to aspects irrelevant to the conversation at hand. Shared meaning is thus still very much of a requirement, and actually expected, given how speakers of the same (linguistic) community are assumed to share *common ground* (Kabbach & Herbelot, 2021, §2.3).

## 3. What exactly is the function of language?

### 3.1. *The need for an alternative hypothesis*

In any case, a characterization of communication success that rests on the hypothesis of shared meaning can only be an approximation at best, for speakers can never actually verify that they do share meaning in practice. As Sperber and Wilson (1986/1995, p.18) emphasize, "when human beings try to communicate with each other, they are aiming at something they can never, in fact, achieve". The

---

[1] And has been extensively discussed in the philosophy of language, notably through the question of meaning holism and instability. See (Jackman, 2020), especially §3.2 and references therein.

formalization of communication success in the code model is psychologically unrealistic, for it would actually require speakers to perform an infinite number of recursive mutual checks on their respective mental content to ensure their making use of the same context and guarantee their encoding and decoding identical messages (Sperber & Wilson, 1986/1995, ch1 §3).

Moreover, if all speakers can do during conversation is strive to *avoid conflict* rather than *enforce agreement*—given that overt disagreement is the only kind of information they actually have access to in practice—the question boils down to whether the *absence of (behavioral) conflict* really equates *(conceptual) agreement*, which is far from obvious. The structure of language and cognition indeed appears to be such that you and I can agree on the fact that *kittens are cute* without having to agree on what kittens are, or up to what age one can call a cat a kitten, on why we find kittens cute, or even on what constitutes a good prototype of cuteness. Many argue that, in such cases, conceptual variability between us does not really matter: language can remain vague and vagueness can even serve communication (van Deemter, 2010). But saying that conceptual variability does not matter does not mean that it is necessarily marginal. Regardless, it remains that language appears to be structurally *underdetermined* regarding concepts, in that my uttering the word *cat* does not even begin to convey the richness of my concept CAT. Therefore, our agreeing on the fact that *kittens are cute* remains what it is: an *appearance* of agreement between what are probably highly subjective and thus clearly distinct conceptual representations, compatible in this context only inasmuch as they do not lead to overt disagreement between us.

Considering subjectivity to be negligible overall in language processing is thus probably somewhat of an arbitrary choice—true *by definition* rather than by empirical observation—and the assumption of shared meaning may very well prove to be biased by our tendency to actually overestimate how much we share meaning with others (Martí, Piantadosi, & Kidd, 2019). Be that as it may, would it really make sense to talk about “communication” if it did not involve shared meaning in the first place? Or if the notion of mutual understanding was not grounded in a form of “sameness of concept” among interlocutors? The language-for-communication hypothesis still resists the relaxing of the shared meaning assumption, even if just for questions of clarity and consistency. If we are to call into question its characterization of success, we thus need more than a new model of communication: we need a new hypothesis about the function of language *itself.*

### 3.2. *The language-for-coordination hypothesis*

The primary contribution of our work is to propose a radical change of perspective on the role subjectivity is assumed to play in language processing (in general) and in interpretation (in particular). In considering that subjectivity is structural rather than marginal, we argue that the function of language is actually best characterized as the coordination rather than the communication between minds.[2] Concretely, we distinguish coordination from communication with respect to their requirement of shared meaning: strict for communication, loose for coordination. In effect, coordination dispenses itself from having to characterize success in terms of necessary conceptual alignment between speakers. Indeed, interpretation can be thought of as a problem of *constraint satisfaction*, but one that admits multiple solutions and that is resolved independently by speakers taken individually, so that they need not converge to identical solutions.

Let us take a concrete example to illustrate our point: it is common in the scientific literature to find counter-arguments to the subjectivity of meaning revolving around the problem of concept-to-word mapping. The argument goes more or less as follows: if you call a *dog* what I call a *cat*, how can the two of us communicate with one-another? Yet, things need not be so binary: there is ample leeway between my saying *cat* when you say *dog* and you and I having *identical* concepts of CAT and DOG. The example is of course limited—what are identical or similar concepts in the first place?—but it gives us an intuitive sense of the claim: the mapping between words and concepts need not be one-to-one, so that different speakers can actually associate different concepts to identical words.

In fact, the process of *making sense* of linguistic signals could perfectly accommodate the pervasiveness of subjectivity specifically because it remains private and because meaning is never negotiated directly across speakers. You and I certainly do have to agree that what we have in front of us is referred to as a *cat* and not a *dog*, but we will never negotiate our conceptual representations of CAT and DOG directly: only the constraints they have to satisfy externally. And since the words *cat* and *dog* are structurally underdetermined with respect to our concepts of CAT and DOG, the problem of constraint satisfaction so defined may very well admit multiple solutions that will be produced privately by each of us, within the subjectivity of our respective conceptual spaces.[3]

---

[2] Our use of the term "coordination" is somewhat arbitrary and primarily motivated by the desire to clearly depart from the term "communication" without resorting to a neologism. Nonetheless, it also rests on the observation that, to coordinate, people need not *do the same thing*. The examples of coordination provided by Clark (1996, p.3) for instance—such as *waltzing* or *playing music*—give a sense of the intuition. From not *doing* the same thing to not *thinking* the same thing, there is a bridge between behavior and mind that we will take the liberty to cross here.

[3] We neither provide nor commit to any particular characterization of the notion of "constraint", though we do commit to the view that concepts are mental entities. In the above example, constraints could for instance be considered "referential", insofar as you and I have to agree that the word *cat* should be used to refer to the animal we have in front of us. But in a different conversation, I may try to convince you that *cats are adorable creatures*, and constraints on usage of *cats* or *adorable* will be formalized differently. Anyhow, the point made by the language-for-coordination hypothesis is that aligning usage does not require aligning concepts. The intuition is plain: we need not agree for the same reasons. "Making sense" only requires that we find at least *one way* to agree, that is, one way to accommodate our subjective concepts to the situation at hand.

## 4. Coordination and subjectivity from an evolutionary perspective

### 4.1. *Language-for-coordination and explanatory adequacy*

Our proposition is not without empirical support. Kabbach and Herbelot (2021), for instance, question the standard view that communication succeeds because the impact of subjectivity is negligible (i.e. the view that speakers of the same linguistic community share significant *common ground* despite having private mental content (Clark, 1996)). They specifically investigate whether common ground emerges from linguistic interactions in the first place, and show using computational models of lexical meaning that aligning different models on *parts* of their semantic spaces does not necessarily lead to increased *overall* similarity between them. Worse, they show that models often manage to improve superficial alignment by actually resorting to idiosyncratic rather than commonly shared aspects of their conceptual spaces, providing thereby a computational characterization of the distinction between agreeing and compatible semantic representations.

Such work, of course, is no definite proof of our argument. But it stands in the long tradition of computational work that, without guaranteeing what *is*, can at least give us some intuition about what *could be*: in this case, superficial behavioral alignment does not have to equate deep conceptual agreement. To us, the role of subjectivity in language should no longer be downplayed: in practice, conceptual similarity across people is extremely hard to validate experimentally (Kabbach & Herbelot, 2021, §2.2) and the fact of the matter is that "different subjects give individually different results on the many tasks about meaning that have been administered over the decades in cognitive psychology" (Pelletier, 2017, p.74). Those empirical data have to be accounted for one way or another.

### 4.2. *Adaptability: the evolutionary benefits of subjectivity*

Interestingly, support for the language-for-communication hypothesis may come from considerations about the emergence of language and asking what benefits subjectivity could have from an evolutionary perspective. One way to concretely tackle the problem is to consider the impact of cumulative culture on adaptability. According to Mithen (1996), human societies distinguish themselves from apes by a striking technological gap, giving our species a major survival advantage as it enables us to adapt much more easily to environmental changes. The superiority of those technological abilities have been said to arise from cumulative culture, i.e., from our ability to innovate by creating new knowledge as well as to preserve existing knowledge by passing it on to other members of the group (Mesoudi, 2011). The question that arises, then, is how to provide a functional characterization of the cognitive processes involved in cumulative culture, and more specifically of the role language plays in both (conceptual) *innovation* and *transfer*.

One interesting take on the matter is the study of Toya and Hashimoto (2018) which investigates the evolutionary benefits of recursion. Drawing a parallel between mental operations and action sequences involved in toolmaking, they show how specific types of recursive operations can lead to improved tool manufacturing strategies and to the production of more and more diverse tools—ultimately leading to improved population fitness, better adaptability and increased survival capabilities. Yet, if their study paves the way for an empirical investigation of the question that interests us here—demonstrating how increased conceptual diversity can positively contribute to innovation and adaptability—it tells us little about *transfer* and how a specific innovation may spread across a whole population. More specifically, will members of a group necessarily transmit *whole* production strategies or rather focus on enabling all members to produce the desired outcome irrespective of the production method used? Which approach will prove better able to entrench innovations and ultimately improve adaptability? This is probably where we can best foresee the possible contribution of the language-for-coordination hypothesis: if what matters is *what* you make and not *how* you make it, relaxing the constraint on shared meaning could prove decisive. As Toya and Hashimoto (2018) indeed detail, a single linearized sequence may correspond to different hierarchical structures (see Figure 1) so that if the point is to match the sequence rather than a particular structure (the *what* versus the *how*) diversity in the *how*—what they call "diversity of production"—can actually prove beneficial. This directly echoes considerations of §3.2 regarding the underdeterminacy of language vis-a-vis concepts and the benefits of subjectivity for coordination.

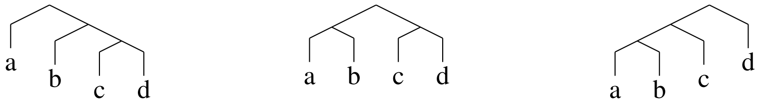


Figure 1.: Different hierarchical structures corresponding to different combinations of recursive operations generating the linearized sequence *abcd*.

We can thus now formulate a clear hypothesis about the possible evolutionary benefits of subjectivity: by enabling agents to coordinate *without* having to align their respective conceptual spaces, language would allow for the spread of a particular conceptual innovation within a group, without losing the benefits for innovation of maintaining distinct conceptual spaces across individuals.

## 5. Conclusion

Is the function of language best characterized in terms of communication or coordination? Inasmuch as communication entails that meaning must be shared across interlocutors when coordination does not, we argue for the latter. Indeed, considering the major evolutionary benefits that subjectivity brings to cumulative culture and adaptability, we argue that we should reconsider the place of subjectivity as being central to any account of human language and cognition, rather than a negligible byproduct of marginally different background experiences grounding conceptual knowledge.

## References

Clark, H. H. (1996). *Using Language.* Cambridge, UK: Cambridge University Press.

Grice, H. P. (1989). *Studies in the Way of Words.* Cambridge, MA: Harvard University Press.

Jackman, H. (2020). Meaning Holism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 ed.). Metaphysics Research Lab, Stanford University.

Kabbach, A., & Herbelot, A. (2021). Avoiding Conflict: When Speaker Coordination Does Not Require Conceptual Agreement. *Frontiers in Artificial Intelligence*, *3*, 95.

Martí, L., Piantadosi, S. T., & Kidd, C. (2019). Same Words, Same Context, Different Meanings: People are unaware their own concepts are not always shared. In *Proceedings of CogSci* (pp. 2296–2302). Montréal.

Medin, D., Goldstone, R., & Gentner, D. (1993). Respects for Similarity. *Psychological Review*, *100*(2), 254–278.

Mesoudi, A. (2011). *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences.* Chicago and London: University of Chicago Press.

Mithen, S. J. (1996). *The Prehistory of the Mind: a Search for the Origins of Art, Religion and Science.* London: Thames and Hudson.

Pelletier, F. J. (2017). Compositionality and Concepts—A Perspective from Formal Semantics and Philosophy of Language. In J. A. Hampton & Y. Winter (Eds.), *Compositionality and Concepts in Linguistics and Psychology* (pp. 31–94). Cham: Springer International Publishing.

Pickering, M. J., & Garrod, S. (2006). Alignment as the Basis for Successful Communication. *Research on Language and Computation*, *4*(2), 203–228.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana, IL: University of Illinois Press.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and Cognition* (Second ed.). Oxford: Basil Blackwell.

Toya, G., & Hashimoto, T. (2018). Recursive Combination Has Adaptability in Diversifiability of Production and Material Culture. *Frontiers in Psychology*, *9*, 1512.

van Deemter, K. (2010). *Not Exactly: In Praise of Vagueness.* Oxford: Oxford University Press.

# LANGUAGE CHANGE IN HINDI FROM A PANINIAN PERSPECTIVE

Aniket Kali*[1] and Gerald Penn[1]

[1]Corresponding Author: aniket.kali@mail.utoronto.ca
[1]Department of Computer Science, University of Toronto, Toronto, Canada

We examine the development of Hindi from Sanskrit using a spectral analysis algorithm. These methods are exceptionally good at drawing out inconsistencies and discrepancies in a dataset, and are therefore well suited to questions around language change. Using a Hindi corpus developed from the Hindi/Urdu Treebank Project (Bhat et al., 2017) annotated in a Paninian (Sanskrit based) scheme, we contrast Panini's *karaka* role scheme against Hindi grammatical markers, such as postpositions. More broadly, we contribute a novel approach to uncovering and analyzing specific changes in language through time.

## 1. Introduction

The most reliable means of documenting language change through time is by directly considering examples drawn from documents distributed over the history of that language. There may, however, be another way. Many languages come with linguistic traditions of their own. If these traditions come to be interpreted prescriptively, the community that participates in the tradition may attempt to accommodate an inherited (semi-)formal theory to their contemporary language, either to demonstrate that the language is still licensed by the tradition, or simply because that is the only alternative that comes to mind for analyzing contemporary language. These accommodations may not be acknowledged or documented, but it may be possible to infer them from example explanations or derivations.

In this paper, we seek to investigate the possibility of such inference to document changes to Hindi syntax relative to the Paninian Sanskrit tradition. Using spectral analysis techniques, we analyze a Hindi corpus that has been annotated according to Paninian conventions by The Hindi/Urdu Treebank Project (Bhat et al., 2017). By treating Hindi as a kind of "noisy" Sanskrit, and imposing a grammar devised for Sanskrit onto it, we show that a simple technique for noise reduction can in fact reveal certain specific changes. In particular, as we describe in detail below, it has been able to reveal changes between the locative and genitive roles and their grammatical markers, as well as a failure to classify the reflexive pronoun *apna*, given the diversity of reflexives in Sanskrit. It also highlights the development of ergativity and Differential Object Marking, both of which developed well after classical Sanskrit.

While numerous mathematical approaches to language change have focused on modelling the system of syntactic change itself (Niyogi & Berwick, 1997; Kodner & Cerezo Falco, 2018), or have sought to understand its cognitive or semantic implications (Hamilton, Leskovec, & Jurafsky, 2016; Habibi, Kemp, & Xu, 2020), to our knowledge, the vast majority have never been used as a means of discovery. Those that have, some of them with very similar methods to the one used here (Belkin & Goldsmith, 2002; Belkin & Niyogi, 2003; Thaine & Penn, 2019) do not rely on a parallel linguistic tradition as a means of identifying changes. We hope this work will contribute to advancing just such an approach.

## 2. Grammar and Corpus

The *Ashtadhyayi*, written c. 350 BCE by Panini, is perhaps the first serious grammar of a language - in its case, Sanskrit. It is admired by linguists of all persuasions for its wide range of descriptive devices, and its degree of formal rigour (Kiparsky, 2009).

One of the work's defining traits is its use of *karaka* roles, abstract case markings with some connection to deep/semantic arguments, at least as they appear in the corpus we used, although the correspondence between theta-roles and *karaka* roles in Panini has been disputed elsewhere (Houben, 1999). Kiparsky (2009) articulates three principles of *karaka* theory: (1) every *karaka* role must be "expressed" by a morphological element, (2) no *karaka* role can be expressed by more than one morphological element, and (3) every morphological element must express something. Hindi is a morphologically poorer language than Sanskrit. As Bharati, Bhatia, Chaitanya, and Sangal (1998) have outlined, this requires an analysis using inflections, postpositions, and auxiliary words rather than one with morphology alone. Bhat et al. (2017) did such an analysis by hand in their corpus.

The Hindi-language fragment of the Hindi/Urdu Treebank Project (Bhat et al., 2017) contains 11,600 sentences and 242,600 tokens. This corpus is annotated using an extension of the Universal Dependencies scheme: roles like 'vmod' have been subdivided into 'k1' (agent), 'k2' (patient), and so on. These 'k*' roles were inspired by traditional, Paninian *karaka* roles. We evaluate the relationship between the actual distribution of Hindi morphemes and clitics and *karaka* roles using spectral analysis.

## 3. Singular Value Decomposition

The corpus annotation scheme of Bhat et al. (2017) uses postpositions to indicate role assignments to syntactic phrases. We define a matrix $\mathbf{A}$ that aligns postpositions with roles, in order to analyze discrepancies in their correspondence. Each row $\mathbf{x}_i$ represents a role, with $n$ total roles. Then row $\mathbf{x}_i = [x_1, ..., x_j, ..., x_m]$ where $x_j$ is the number of times the $j$-th postposition is the final element in a phrase marked with the $i$-th role, with $m$ total postpositions.

We extracted a list of roles and postpositions from the corpus to construct this matrix. We only counted postpositions at the end of phrases, as these determine the role, and counted phrases without postpositions as inflected with a NULL postposition. In the case of inflection by gender and number, e.g., the possessive *ka*, which appears as *ki* (singular or plural feminine) or *ke* (plural masculine), we count it as the stem *ka*. This provides consistency, as gender and number are determined by the possessee and have no bearing on the role itself. Nonetheless, we analyze occurrences of the form *ke*+PSP as is (where PSP is the POS tag for postposition), since this is a postpositional construct in Hindi that is unrelated to the possessive. In the case of pronouns, Hindi also permits postpositions to attach as morphology, which we extract and analyze in the manner described above.

In our presentation, we retain all roles found in the corpus, but include only the top 26 postpositions, to reduce clutter. These make up about 99.9% by occurrence. In practice, including the rest has a negligible effect on our analysis. This leaves us with a total of $n$=65 roles and $m$=26 postpositions. A complete list of roles and postpositions, and their frequencies, are found in the supplementary materials.[1]

Singular-Value Decomposition (SVD) takes a co-occurrence matrix as input and returns component vectors that capture the most salient dimensions of the dataset, as measured by variance, together with weights, called singular values, that describe the importance of each dimension. Here, we present only the dimensions corresponding to the six largest singular values and project the vectorized roles onto them pairwise in a series of two-dimensional visualizations. We used `scikit-learn`'s implementation of SVD.

The $j$-th component in a vector represents the relevance of the $j$-th postposition. Once normalized, entries with an absolute value close to 1 are the most relevant, and those close to 0, irrelevant. Negative entries indicate an anti-correlation, which can still be very relevant.

Examining the projection of roles onto components outlines the relationship between particular roles and postpositions. Components with a single dominant postposition should attract roles highly related to the functional use of that postposition. Likewise, components containing multiple postpositions may either attract multiple roles, or a single multi-purposed role, indicating an analytical gap.

## 4. Results

As stated above, the most relevant parts of a component are those entries with values furthest from 0. We provide the first three postpositions ranked this way, in Table 1. These entries already tend rapidly toward 0, which is where the vast majority of entries lie anyhow. As we will explain, there are particular reasons that these postpositions appear on the components that they do, and that certain roles associate with them.

---

[1] https://hdl.handle.net/1807/117650

Table 1. The first six components from our spectral analysis. The three most extremal postpositions of each component are given, as determined by the absolute value of their singular value.

| Component | Postposition 1 | Postposition 2 | Postposition 3 |
|---|---|---|---|
| $c_0$ | *ka* (0.97) | NULL (0.21) | *ne* (0.06) |
| $c_1$ | NULL (0.9) | *ne* (0.32) | *ka* (-0.22) |
| $c_2$ | *mein* (0.92) | *par* (0.37) | *tak* (0.01) |
| $c_3$ | *ne* (0.86) | *ko* (-0.41) | NULL (-0.24) |
| $c_4$ | *ko* (-0.88) | *ne* (-0.36) | NULL (0.25) |
| $c_5$ | *ke liye* (0.99) | *se* (-0.11) | *ke karan* (0.0) |

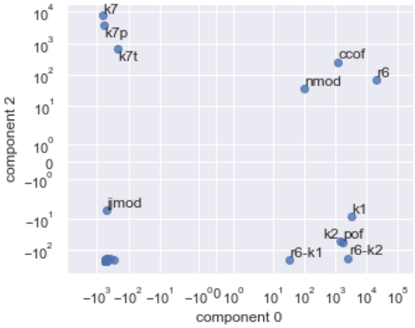


Figure 1. Projection onto components 0 and 2.

We consider roles mapped onto the components $c_0$ and $c_2$ first, shown in Figure 1. Each dot is a role, where the most relevant roles are those furthest away from the main cluster. We do not annotate the main cluster for this reason in any figure, although we do provide a list of cluster roles in the supplementary materials. It is worth noting that frequency plays a part in where roles land: a very frequent role can end up seeming more relevant than it is, with a weak correlation exaggerated by its prevalence. On the other hand, SVD remains exceptionally good at highlighting lower frequency roles that are actually relevant.

Considering $c_2$ first, the roles k7t (location in time), k7p (location in space) and k7 (location elsewhere) as the most prominent. Given that $c_2$ is mainly composed of *mein* (in) and *par* (on), this makes total sense: the locative postpositions relate to the locative roles. Considering the other five components, it is striking that neither *mein* nor *par* are relevant in any. Nor, as we will see, are the locative roles as a group so clearly related to any other component. While this is a more straightforward case, it also reinforces the empirical validity of this analysis.

We consider $c_0$ next, where the role r6 (genitive) and its variant r6-k2 (complex predicate with object) dominate, accompanied by k1 (agent), likely due to sheer frequency. Notice that $c_0$ is composed primarily of *ka* (possessive marker), along with NULL, albeit more weakly. The relationship between *ka* and r6 is clear, with the effect of relating the possessive marker to the genitive role.

r6 appears with the NULL marker in the corpus almost exclusively with the reflexive pronoun *apna*, which is marked in the corpus as r6, as are other possessive pronouns. The inchoate category of “reflexive pronoun” only entered linguistics in the late 19th century, partly due to comparative evidence from Sanskrit (Orqueda, 2019). Other possessive determiners do exist in Sanskrit, although historically it has always been more common to use the genitive case of the corresponding pronoun, which thus exhibits no agreement with its head noun. The reflexives stand out in Sanskrit by being so heterogeneous in any account, including the Paninian one, as they include an indeclinable (*swayam*), a noun (*atman*), a declinable determiner (*svah*) that, unlike every other possessive declinable determiner in Sanskrit, is not depronominal, and in certain cases, the verbal middle voice (known to the traditional grammarians as *atmanepada*, arguably translatable as the reflexive voice). Against this backdrop, just about any choice of label for Hindi’s *apna* would be a stretch, and indeed the r6 label belies the fact that *apna* is declinable and, if it is de(pro)nominal at all, not transparently so.

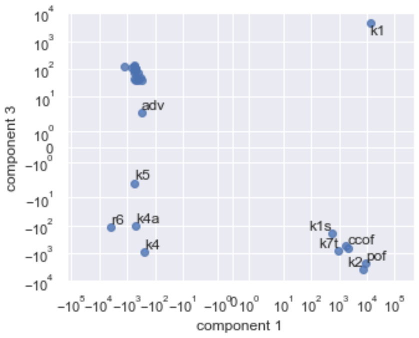


Figure 2. Projection onto components 1 and 3

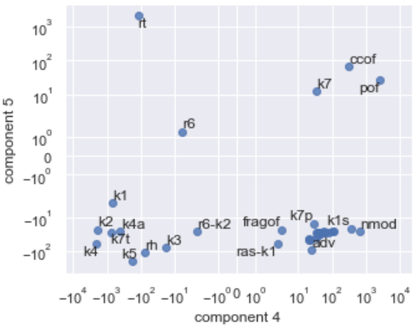


Figure 3. Projection onto components 4 and 5

Before we discuss the next group of components, we find it necessary to provide some background on ergativity in modern Hindi. In Hindi, the subject of an intransitive clause and the object of a transitive clause are treated the same way, whereas the subject of a transitive clause is marked with *ne*. But Hindi is split-ergative: we only see the ergative pattern in the preterite and perfect tenses (Verbeke & De Cuypere, 2009). The origin of this construction is an open question with differing hypotheses, but crucially, it lacks any precedent in Sanskrit’s case system itself (Anderson, 1977; Butt & King, 2004; Verbeke & De Cuypere, 2009), presenting a problem for annotating Hindi using a Paninian scheme.

As for components $c_1$, $c_3$, and $c_4$ (Figures 2 and 3), what is immediately clear is that there exists some relationship between NULL, *ne* (ergative marker), and *ko* (to). These postpositions are almost entirely isolated to within this component group, with the exception of NULL in $c_0$, which we have already addressed. Looking at Figure 2, k1 and k2 (patient) are consistently prominent throughout. This is rather conspicuous, as *ne* almost exclusively occurs with k1 to mark ergative agents, and *ko*, primarily with k2 but rarely k1. NULL however occurs frequently with both roles, and must therefore logically serve to unify them, although this alone does not explain the linguistic relationship between *ne* and *ko*.

We believe the solution comes through a phenomenon known as Differential Object Marking (DOM). DOM is a process of marking animacy and definiteness in a language. It is present in Hindi, and developed alongside the modern language - well after Sanskrit. In Hindi, its presence is typically expressed with *ko*, and its absence, with NULL (Montaut, 2018). This latter point is crucial, as we can relate it to ergativity: NULL marks both non-ergative agents, and non-animate or non-definite objects. In other words, the development of ergativity and DOM in Hindi - phenomena both not present in Sanskrit - have come to intersect through NULL.

The component $c_4$ also provides brief additional insight into role mergers connected with *ko*. As Butt and King (2004) note, *ko* is a postposition of many forms, covering the accusative (including DOM as above), dative, and experiencer or dative passive. This is precisely what we observe in Figure 2, with k2, k4 (recipient), and k4a (experiencer), respectively present on the negative side of $c_4$ where *ko* has a strongly negative singular value. Additionally, k7t appears in this group, because temporal locatives in Hindi can also be marked with *ko*.

Finally, we examine $c_5$ in Figure 3. The postposition *ke liye* (for) almost always occurs under rt (purpose), whereas *se* (from) occurs primarily with k5 (source). The relationship between these two is rather one-sided: *se* is a rather versatile postposition, and occasionally occurs under rt, but we never observe *ke liye* as expressing k5. *se*, moreover, is actually the third most frequent postposition on the role rt, after *ko*. A *post hoc* analysis of the corpus reveals that under the role rt both postpositions are used to mark adverbials in Hindi, demonstrating a convergence of one aspect of *se* with *ke liye* in this domain.

The reader familiar with Universal Dependencies will notice that we have not commented on nmod (noun participle), pof (part of relation), or ccof (coordination or conjunct), despite their prevalence. This is because in this corpus they either serve as catch-all roles, or annotate more functional aspects of grammar. They have less to do with the *karaka* scheme (Begum et al., 2008; Tandon, Chaudhry, Bhat, & Sharma, 2016), and in turn the aim of this work.

## 5. Discussion

All three of our findings can be corroborated by existing knowledge among area specialists in Hindi syntactic innovations. This suggests that spectral analysis relative to a prescriptive grammatical structure in which a descendant language is viewed as a noisy version of its ancestor has some promise as a means of discovering as-yet unknown changes.

The success of this work has relied on the alignment of several stars. First, we picked a language pair with a very mature and well-documented linguistic tradition; Sanskrit is the most analyzed language in the world after only modern English. Second, the Hindi/Urdu Treebank Project methodically and, to our great relief, electronically annotated enough Hindi data that SVD could reveal certain anomalies. But most importantly, they did so in accordance with certain aspects of the ancient Paninian tradition that made the application of our proposed method straightforward. There are other approaches to Hindi grammatical analysis. Current analyses of modern Hindi are rooted in Prakritic grammars dating back to 1698 at the earliest (Bhatia, 1987), but these grammars come from a colonial tradition based in the Western classical paradigms of Greek and Latin.

It is an open question as to whether the method proposed here can be extended to pairs of cognate languages other than direct ancestor-descendant pairs. There are a variety of more sophisticated methods for spectral analysis than SVD, the algorithms for which were worked out in the 1960s. Some of them would perhaps reveal anomalies between more diverse pairs, or more subtle anomalies with ancestors and descendants. Having located an anomaly, its interpretation furthermore remains non-trivial. We relied on the assistance of Professor Miriam Butt, an area specialist, who was immediately able to interpret what the anomalies were. The present authors would not have been able to do this themselves, even with considerable effort.

It should also be noted that we did not uncover every change from Sanskrit to Hindi, nor even every noteworthy change by conservative estimates, such as the development of the dative passive in Hindi. All of these shortcomings remain areas for further investigation.

## Acknowledgements

Our many thanks to Professors Miriam Butt and Dipti Sharma for their input and feedback on the development of this work, and the reviewers for their comments.

## References

Anderson, S. (1977). On mechanisms by which languages become ergative. *Mechanisms of Syntactic Change*, 317-363.

Begum, R., Husain, S., Dhwaj, A., Sharma, D. M., Bai, L., & Sangal, R. (2008). Dependency annotation scheme for Indian languages. In *Proceedings of the Third International Joint Conference on Natural Language Processing* (p. 722-726).

Belkin, M., & Goldsmith, J. (2002). Using eigenvectors of the bigram graph to infer morpheme identity. In *Proceedings of the ACL-02 Workshop on Morphological and Phonological Learning* (pp. 41–47).

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, *15*(6), 1373-1396.

Bharati, A., Bhatia, M., Chaitanya, V., & Sangal, R. (1998). Paninian grammar framework applied to English. In *South Asian Language Review* (Vol. 8). Creative Books.

Bhat, R., Bhatt, R., Farudi, A., Klassen, P., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D. M., Vaidya, A., Vishnu, S. R., & Xia, F. (2017). The Hindi/Urdu treebank project. In *Handbook of Linguistic Annotation* (pp. 659–697). Springer.

Bhatia, T. (1987). *A history of the Hindi grammatical tradition: Hindi-Hindustani grammar, grammarians, history and problems.* Brill.

Butt, M., & King, T. (2004). The status of case. In V. Dayal & A. Mahajan (Eds.), *Clause Structure in South Asian Languages* (p. 153-198). Springer.

Habibi, A., Kemp, C., & Xu, Y. (2020). Chaining and the growth of linguistic categories. *Cognition*, *202*(104323).

Hamilton, W., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In *Proc. 54th ACL* (pp. 1489–1501).

Houben, J. (1999). 'Meaning statements' in Panini's grammar: on the purpose and context of the Ashtadyayi. In *Studien zur Indologie und Iranistik* (Vol. 22, p. 23-54).

Kiparsky, P. (2009). On the architecture of Panini's grammar. In *Sanskrit Computational Linguistics* (p. 33–94). Springer.

Kodner, J., & Cerezo Falco, C. (2018). A framework for representing language acquisition in a population setting. In *Proc. 56th ACL* (pp. 1149–1159).

Montaut, A. (2018). The rise of differential object marking in Hindi and related languages. In *Diachrony of differential argument marking* (p. 281–313). Language Science Press.

Niyogi, P., & Berwick, R. (1997). A dynamical systems model for language change. *Complex Systems*, *11*(3).

Orqueda, V. (2019). *Reflexivity in Vedic.* Brill.

Tandon, J., Chaudhry, H., Bhat, R., & Sharma, D. M. (2016). Conversion from Paninian karakas to universal dependencies for Hindi dependency treebank. In *Proc. 10th Linguistic Annotation Workshop, 54th ACL* (pp. 141–150).

Thaine, P., & Penn, G. (2019). Vocalic and consonantal grapheme classification through spectral decomposition. In Y. Haralambous (Ed.), *Graphemics in the 21st century* (pp. 367–386). Fluxus.

Verbeke, S., & De Cuypere, L. (2009). The rise of ergativity in Hindi: Assessing the role of grammaticalization. *Folia Linguistica Historica*, *43*, 367-389.

# EMERGENCE OF NEGATION IN A TURKISH HOMESIGN SYSTEM: INSIGHTS FROM THE FAMILY CONTEXT

UĞUR KAN*[1,3], KADİR GÖKGÖZ[2], BEYZA SÜMER[3,5], ELVAN TAMYÜREK[2],
and ASLI ÖZYÜREK[3,4]

*Corresponding Author: ugur.kan@boun.edu.tr
[1]Cognitive Science, Boğaziçi University, İstanbul, Turkey
[2]Linguistics, Boğaziçi University, İstanbul, Turkey
[3]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[4]Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, the Netherlands
[5]Amsterdam Center for Language and Communication, University of Amsterdam, the Netherlands

Deaf children without access to conventional sign language develop gesture systems called homesigns to communicate with their immediate family. Despite the lack of a language model, these systems exhibit many properties of natural languages, including basic syntax [1], stable lexicon [2], complex sentences, and noun phrases [3-4]. One of the challenges in such language creation situations is understanding whether and how the gestures of the family members play a role in the creation of homesign systems. Despite acknowledging the complexity of the gestures in the input, previous studies did not conduct a systematic or detailed analysis of gestures used by different caregivers in the family and their potential impact on homesigners' gestures [5-6]. Thus to what extent gestures used by family members surrounding a homesigner look analogous to homesigners' gestures is understudied. Here, our study focused on a Turkish homesigner child and his family to examine the role of the language model in the emergence of negation, a universal component of human language.

Previously, Franklin et al. [7] analyzed the negation patterns of an American homesigning child, David, on eight play sessions beginning from age 2;10 to 3;11 (years; months). They found that David predominantly used side-to-side headshakes as the main negation marker (84% of 327 negative sentences). Researchers assumed that he co-opted this negative marker from the surrounding hearing culture because his mother and siblings have never been exposed to sign language. Interestingly, the researchers concluded that his negated expressions reflected the ability of the child to re-invent negation marker without an available language model to guide him. However, in their study, gestures used by different family members around the child were not analyzed to see whether they set up a consistent negation model for him. Thus, the effect of the complexity of the input from language models in the homesign system remains partially unanswered.

To address this issue, we investigate whether the hearing family members of a Turkish homesigner child display a rich and consistent negation model with their gestures, and the child benefits from this model. We observed natural play sessions (182 minutes) of a Turkish deaf child who had severe hearing loss playing with his hearing mother and 13-years-old sister during six different time points beginning from the child's age 5;11 until 6;3. Since there is no comprehensive study on the co-speech gestures of Turkish speakers, the coding template consisted of the negation markers of the Turkish Sign Language [8-9]. The child's and family members' gestures were coded for the presence and forms of negative markers only and compared to each other in terms of form and frequency. Table 1 (see supplementary materials) presents preliminary results of negation forms used by each family member. The child produced side-to-side headshakes as the most frequent marker (39% of all his negative sentences produced). Nevertheless, headshakes were not as common among his mother and sister's negation forms (4,5% and 10%, respectively). Contrary to the child, his mother produced a backward head tilt as the most frequent marker (66% of all her negative sentences), while his sister used a 1-handshape variant (60% of all her negative sentences). They both seemed to present consistent negative strategies as a model and differ from each other and the child. Although their presence in the play sessions varied, the mother and the sister provided significant input (44 and 30 negation gestures in 85 and 97 min. they interacted with the child in sessions, respectively). In comparison, the child produced 102 negation gestures in 182 minutes.

These preliminary results suggest that homesigners can receive rich environmental input [10-12]. However, this input might not be consistent between members of the family. The fact that we found a variation among family members in the types of negation markers is in line with recent findings showing that languages of small communities exhibit more variation than in larger ones [13]. These results also reinforce the previous observation that homesigners are not getting their system from the input [14]: 39% of the child's negative sentences use side-to-side headshake, which does not appear to be the most frequent strategy of either his mother or sister. Finally, the child's most frequent marker supports previous claims that some properties of language are resilient and are less inclined to be affected by input during the creation [1], even when the input can be rich. Rather, the resemblance of reliance on side-to-side headshake between the child and David also suggests a universally emerging negation strategy. Overall, the study aims to contribute to the scarce literature on how one of the main characteristics of human language, negation, can emerge in a homesign environment where hearing and deaf interlocutors actively participate. Further functional analyses of the negation markers in the gesture utterances of the interactants will be made.

## References

[1] Goldin-Meadow, S. (2003). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. Psychology Press. Berlin, Boston: De Gruyter Mouton.

[2] Richie, R., Yang, C., & Coppola, M. (2014). Modeling the emergence of lexicons in homesign systems. Topics in cognitive science, 6(1), 183-195.

[3] Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, *391*(6664), 279-281.

[4] Hunsicker, D., & Goldin-Meadow, S. (2012). Hierarchical structure in a self-created communication system: Building nominal constituents in homesign. Language, 88(4), 732.

[5] Goldin-Meadow, S., Mylander, C., de Villiers, J., Bates, E., & Volterra, V. (1984). Gestural communication in deaf children: The effects and noneffects of parental input on early language development. Monographs of the society for research in child development, 1-151.

[6] Flaherty, M., Hunsicker, D., & Goldin-Meadow, S. (2021). Structural biases that children bring to language learning: A cross-cultural look at gestural input to homesign. *Cognition*, b*211*, 104608.

[7] Franklin, A., Giannakidou, A., & Goldin-Meadow, S. (2011). Negation, questions, and structure building in a homesign system. Cognition, 118(3), 398-416.

[8] Zeshan, U. (2006). Negative and interrogative structures in Turkish Sign Language (TID). Interrogative and negative constructions in sign languages, 128-164.

[9] Gökgöz, K. (2011). Negation in Turkish Sign Language: The syntax of nonmanual markers. *Sign Language & Linguistics*, *14*(1), 49-75.

[10] Haviland, J. (2020). Zinacantec family homesign (or “Z”). In O. Le Guen, J. Safar & M. Coppola (Ed.), Emerging Sign Languages of the Americas (pp. 393-400).

[11] Mesh, K., & Hou, L. (2018). Negation in San Juan Quiahije Chatino Sign Language: The integration and adaptation of conventional gestures. Gesture, 17(3), 330-374.

[12] Reed, L. (2021). Sign networks: Nucleated network sign languages and rural homesign in Papua New Guinea. Language in Society, 1-35.

[13] Raviv, L., Meyer, A., & Lev-Ari, S. (2020). The role of social network structure in the emergence of linguistic structure. *Cognitive Science*, *44*(8), e12876.

[14] Carrigan, E., & Coppola, M. (2012). Mothers do not drive structure in adult homesign systems: Evidence from comprehension. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).

# EVOLVING LINGUISTIC DIVERGENCE IN SOCIO-POLITICAL POLARITIES

Andres Karjus*[1] and Christine Cuskley[2]

*Corresponding Author: akarjus@tlu.ee
[1]ERA Chair for Cultural Data Analytics, Tallinn University, Tallinn, Estonia
[2]School of English Literature, Language and Linguistics, Newcastle University, Newcastle upon Tyne, United Kingdom

## 1. Introduction

All human languages change over time, as linguistic variants are discarded, innovated, and their meanings shifted. Most change stems from variation, be it geographical, cultural or social. Here we examine a division and source of variation intersecting these categories: political polarization. Particularly in the last decade, social and political scientists have been concerned with the causes and alarming social effects of increasing media polarization and partisan segregation particularly in the US (Brown & Enos, 2021). However, political polarization may also have an effect on the dynamics of language evolution and change, forming the basis for signals of in-group and out-group status (?) and potentially leading to more dramatic language speciation at deeper timescales (?; Altmann, Pierrehumbert, & Motter, 2011). Using existing political categorizations, we collect and use an extensive corpus of social media posts to quantify divergence in American English along the left-right axis, both in topics of conversation and lexical semantics.

Twitter data has been shown to be useful for mapping lexical innovation and variation (Grieve, Nini, & Guo, 2018; Bhat & Klein, 2020) and analyzing polarization effects (Chen, Salloum, Gronow, Ylä-Anttila, & Kivelä, 2021). Studies of linguistic divergence between political divides have often focused on politicians and activists (Adamic & Glance, 2005; Li, Schloss, & Follmer, 2017; Gentzkow, Shapiro, & Taddy, 2019). Those on the general population have grouped subjects based on self-reporting (Halpern & Rodrıguez, 2018) and social media activity (Sylwester & Purver, 2015; Demszky et al., 2019; KhudaBukhsh, Sarkar, Kamlet, & Mitchell, 2020). We focus on everyday interactions, not just political communication, and scale up the latter sampling approach, mining and grouping a total of half a billion follower listings across 72 large US news media accounts. We use the Allsides Media Bias Rankings (Allsides, 2021, v4) to delineate likely left and likely right biased news outlets (Allsides is not perfectly unbiased itself, but serves as a useful starting point). Having access to the entire follower bases of the news accounts allows us to carry out full set operations, and find users who only follow one side but not the other. Further limiting these to active and identifiably US-located users (a major bottleneck) leaves 6202 likely "left" and 4783 likely "right" aligned accounts. We mined their tweets between February-September 2021, yielding a corpus of 1.5 million tweets (750,814 and 732,521, respectively).

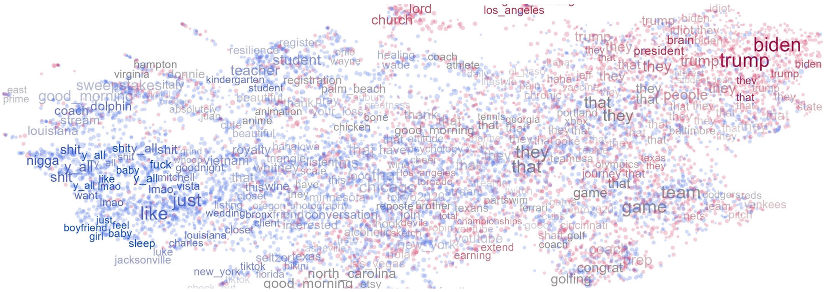


Figure 1. Divided they (micro)blog. UMAP dimension reduction of an LDA topic model of the tweet corpus, with illustrating keywords. Each dot is an account, accounts with similar content are arranged together (blue = "left", red = "right").

## 2. Results and Conclusions

We fit a topic model to the tweet corpus, visualized in Figure 1, showing how the two sides differ in their daily conversation topics. Word frequencies quantify this further, revealing magnitudes of difference in the usage of some terms. Phrases like *President Trump* and *communist* are used about 16x more by the "right" in 2021, who also prefer hand-shaped emoji in contrast to some face-shaped emoji used predominantly by the "left". We also apply word embeddings to detect semantic change and identify potentially competing variants. While users on either side still of course speak the same language, a number of interesting divergences emerge, with differences in the meaning of some emoji and certain politically charged terms (e.g. *woke* referring to either "wokeness" or just waking up). Ongoing research aims to validate these findings with a crowd-sourced annotation task (Schlechtweg, Walde, & Eckmann, 2018), and compare with recent large-scale dialectal work (Grieve et al., 2018) to tease apart sources of variation.

By building on and scaling up previous methodologies of sampling utterances by speaker political alignment, our results show that existing political divisions are already being reflected in language use. Mapping this ongoing lexical and semantic evolution in American English provides a model applicable to studying similar phenomena in any language where sufficient data can be acquired. We emphasize the potential for studies of this kind to shed light on the interplay of interacting evolutionary dynamics at socio-cultural and linguistic levels.

## References

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. In *Proceedings of the 3rd international workshop on Link discovery* (pp. 36–43). Association for Computing Machinery.

Allsides. (2021). *Allsides Media Bias Ratings, version 4.* Available from: https://www.allsides.com/media-bias/media-bias-ratings.

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PLOS ONE*, *6*(5), 1–12.

Bhat, P., & Klein, O. (2020). Covert Hate Speech: White Nationalists and Dog Whistle Communication on Twitter. In G. Bouvier & J. E. Rosenbaum (Eds.), *Twitter, the Public Sphere, and the Chaos of Online Deliberation* (pp. 151–172). Springer International Publishing.

Brown, J. R., & Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour*, *5*(8), 998–1008.

Chen, T. H. Y., Salloum, A., Gronow, A., Ylä-Anttila, T., & Kivelä, M. (2021). Polarization of climate politics results from partisan sorting: Evidence from Finnish Twittersphere. *Global Environmental Change*, *71*, 102348.

Demszky, D., Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., & Jurafsky, D. (2019). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv: 1904.01596.*

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica*, *87*(4), 1307–1340.

Grieve, J., Nini, A., & Guo, D. (2018). Mapping lexical innovation on American social media. *Journal of English Linguistics*, *46*(4), 293–319.

Halpern, D. J., & Rodrıguez, P. L. (2018). Partisan Representations: Partisan Differences in Semantic Representations and their Role in Attitude Judgments. *Proceedings of CogSci 2018*, 6.

KhudaBukhsh, A. R., Sarkar, R., Kamlet, M. S., & Mitchell, T. M. (2020). We don't speak the same language: Interpreting polarization through machine translation. *arXiv: 2010.02339.*

Li, P., Schloss, B., & Follmer, D. J. (2017). Speaking two "Languages" in America: A semantic space analysis of how presidential candidates and their supporters represent abstract political concepts differently. *Behavior Research Methods*, *49*(5), 1668–1685.

Schlechtweg, D., Walde, S. Schulte im, & Eckmann, S. (2018). Diachronic usage relatedness (DURel): A framework for the annotation of lexical semantic change. In *Proceedings of NAACL 2018* (pp. 169–174).

Sylwester, K., & Purver, M. (2015). Twitter Language Use Reflects Psychological Differences between Democrats and Republicans. *PLOS ONE*, *10*(9), e0137422.

# BEHAVIORAL EVIDENCE FOR REPRESENTATIONS OF THE OTHER'S ACTION DURING A JOINT TASK IN RATS

NORIKO KATSU*[1, 2], and KAZUO OKANOYA[1,3]

*Corresponding Author: katsuno0414@gmail.com
[1] Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
[2] Graduate School of Human Sciences, Osaka University, Osaka, Japan
[3] Teikyo University Advanced Comprehensive Research Organization, Tokyo, Japan

## 1. Introduction

Shared intentionality is one of the essential basis for linguistic communication in humans (Tomasello, Carpenter, Call, Behne, & Moll, 2005). It refers to the motivation to share a psychological state and collaborate with others for shared goals and actions. Various cooperative behaviors have been reported in nonhuman species including rats (*Rattus norvegicus*). Nevertheless, it is unclear whether such behaviors are established based on each individual's coordination or independent contributions. Therefore, we used the Joint Simon task to examine whether the rats have representation for other's action to the shared goal.

The joint Simon task is based on the Simon task. In an auditory discrimination task in which correct response is a left lever to stimulus A, and a right lever to stimulus B for example, the response is faster and more accurate when the A is presented from the left than right side. This is because the stimulus and the response are incompatible in the latter condition (Simon effect). Generally, this effect disappears when the task is divided into left and right (half task) but reappear when this task is performed by two people (joint Simon task). This joint Simon task is used to examine the shared representation of the action of a partner during joint activity in humans (Sebanz, Knoblich, & Prinz, 2003).

A previous study demonstrated the joint Simon effect in common marmosets between familiar pairs (Miss & Burkart, 2018). However, the degree of representation sharing may differ according to familiarity between partners, as in humans. Therefore, we also compared the effects based on familiarity with the partner; cagemate and non-cagemate pairs. Subjects were also assigned to mixed and single strain cagemate conditions to manipulate familiarity between strains.

## 2. Methods

The subjects were eight male rats (four Long-Evans and four Wistar strain), housed in pairs. We used an operant box that could be divided into two chambers using a wire-mesh wall. Rats were trained individually on a two-choice auditory discrimination task. All subjects learned to respond by pressing the left lever at 2 kHz, and the right lever at 4 kHz. Unlike in the training, the stimuli were presented from either the left or right in the tests. The subjects experienced the four task conditions; a single condition in which one individual performed a full Simon task, a control condition in which one individual performed a half task, a joint condition in which two individuals shared the left and right half tasks, and a paired control condition in which two individuals were in the operant box but only one performed the half task. The subjects were also tested for joint and paired control tasks both with a cagemate and a non-cagemate. The compatibility effect was used for the index of the Simon effect, calculated by subtracting the correct rate for incompatible from compatible trials. The effect of task condition and familiarity on the compatibility effect was examined by linear mixed models (LMMs).

## 3. Results and Discussion

The subjects showed larger compatibility effects in the single and joint conditions than in control or paired control conditions. Overall, the rats showed both Simon and joint Simon effects. The difference between the paired control and joint conditions indicates that this effect was caused by sharing the task, not merely due to existence of the partner. The effects were significantly larger in cagemate than in non-cagemate pairs. However, no significant differences were found based on whether the partner was of the same or different strain. Therefore, familiarity between partners only partially explain the differences in compatibility effects. In summary, action co-representation for partner in joint tasks was shown in rats, which is highly social species. Although cooperative tendency differed based on familiarity of the strain in the previous study, it is unclear whether social factors modulated the joint Simon effect in this study. Therefore, further studies focusing on individual differences in the effect, such as the frequency of paying attention to the partner, would be informative.

## Acknowledgements

This study was supported by JSPS KAKENHI (No. 19K20643 and No. 21J40080) to NK, and MEXT/JSPS KAKENHI (#4903, No. JP17H06380) to KO.

## References

Miss, F. M., & Burkart, J. M. (2018). Corepresentation During Joint Action in Marmoset Monkeys (*Callithrix jacchus*). *Psychological Science, 29*, 984-995. doi:10.1177/0956797618772046

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition, 88*, B11-B21. doi:10.1016/s0010-0277(03)00043-x

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675-691; discussion 691-735. doi:10.1017/S0140525X05000129

# A CRITICAL POPULATION THRESHOLD FOR CONTACT-INDUCED SIMPLIFICATION

HENRI KAUHANEN*[1]

*Corresponding Author: henri.kauhanen@uni-konstanz.de
[1]Department of Linguistics, University of Konstanz, Konstanz, Germany

Learning biases have long been theorized to play a causal role in the cultural evolution of linguistic systems. In particular, existing literature suggests that the difficulty encountered by second-language (L2) learners in acquiring complex linguistic features may contribute to the loss of those features from the target language in situations of language contact (Bentz & Winter, 2013; Berdicevskis & Semenuks, 2022; Lupyan & Dale, 2010; Sinnemäki & Di Garbo, 2018; Trudgill, 2011; Walkden & Breitbarth, 2019; Weerman, 1993; also see Jansson, Parkvall, & Strimling, 2015 on modelling creolization). Against the backdrop of this body of research, it is reasonable to expect that the population fraction of L2 learners may act as a bifurcation parameter: if sufficiently many L2 learners are present in a speech community, the loss of L2-difficult features may be permanent. Without an explicit model combining population and learning dynamics, however, it is impossible to say where the critical value of such a putative bifurcation parameter might lie.

We propose such a model by extending the variational learner (Yang, 2002) to cover L2 as well as L1 acquisition. For L2 (but not L1) learners, the extended model includes a learning bias that works against the successful (native-like) acquisition of the L2-difficult variant. The asymptotic dynamics of this extended learning model can be studied just like those of the ordinary linear reward–penalty learning scheme (Bush & Mosteller, 1955) that underlies the variational learner. In particular, we show that an L2 learner's expected probability of employing an L2-difficult grammar $G_1$ over its easier-to-acquire competitor $G_2$ tends to a definite value as learning iteration tends to infinity.

Taking the usual infinite learner limit (cf. Yang, 2000) then yields a deterministic dynamical system that describes the evolution of a mixed population of L1 and L2 speakers. This system has three parameters: $\sigma$, the fraction of L2 speakers in the population; $D$, the learning-theoretic strength of the L2-difficulty of $G_1$; and $\alpha$, the fitness ratio (Kauhanen & Walkden, 2018) of the two grammars.

We show analytically that this system always has exactly one stable equilibrium. The system's dynamics are, however, separated into two phases: in one phase, the stable equilibrium satisfies $p > 0$ and $q > 0$, where $p$ and $q$ stand for the probability of the L2-difficult grammar $G_1$ in the L1 and L2 populations, respectively. In other words, the L2-difficult grammar is retained in each population at some non-zero (and possibly high) frequency. In the second phase, however, the attractor is the origin $(p, q) = (0, 0)$, meaning that the L2-difficult grammar is wiped out from *both* populations, including the L1 speaker population which itself is not subject to the learning bias (but feels its effects through interactions with the L2 population). This bifurcation occurs as $\sigma$ crosses the critical value

$$\sigma_{\text{crit}} = \frac{(\alpha - 1)(D + 1)}{\alpha D}, \tag{1}$$

that is, fractions of L2 speakers $\sigma > \sigma_{\text{crit}}$ exhibit simplification dynamics (Fig. 1).

To provide some empirical support for the model, we estimate the parameters $\sigma$ and $\alpha$ from demographic and corpus data, and provide reasonable orders of magnitude for the learning bias $D$, for two historical developments: the loss of verbal inflection in Afrikaans (Trudgill, 2011) and the partial loss of null subjects in Afro-Peruvian Spanish (Sessarego & Gutiérrez-Rexach, 2018). Empirically, the simplification process in Afrikaans went to completion, whereas in Afro-Peruvian Spanish null subjects retain a partial status. These facts are predicted by the model, in the sense that $\sigma > \sigma_{\text{crit}}$ in the former case but not in the latter.

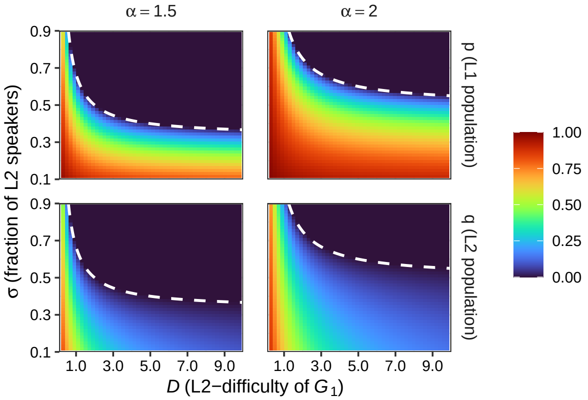


Figure 1. Stable equilibrium $(p, q)$ of the mixed speech community (top row: probability of $G_1$ in L1 speakers; bottom row: L2 speakers). Full simplification occurs above the bifurcation threshold $\sigma_{\text{crit}}$ (equation 1), depicted as the dashed white curve.

## Acknowledgements

Funded by the European Research Council as part of project STARFISH (851423).

## References

Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, *3*, 1–27.

Berdicevskis, A., & Semenuks, A. (2022). Imperfect language learning reduces morphological overspecification: experimental evidence. *PLoS ONE*, *17*(1), e0262876.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning.* New York, NY: Wiley.

Jansson, F., Parkvall, M., & Strimling, P. (2015). Modeling the evolution of creoles. *Language Dynamics and Change*, *5*(1), 1–51.

Kauhanen, H., & Walkden, G. (2018). Deriving the constant rate effect. *Natural Language & Linguistic Theory*, *36*(2), 483–521.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, *5*(1), e8559.

Sessarego, S., & Gutiérrez-Rexach, J. (2018). Afro-Hispanic contact varieties at the syntax/pragmatics interface: pro-drop phenomena in Chinchano Spanish. In J. King & S. Sessarego (Eds.), *Language variation and contact-induced change: Spanish across space and time.* Amsterdam: Benjamins.

Sinnemäki, K., & Di Garbo, F. (2018). Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: a typological study of verbal and nominal complexity. *Frontiers in Psychology*, *9*, 1141.

Trudgill, P. (2011). *Sociolinguistic typology: social determinants of linguistic complexity.* Oxford: Oxford University Press.

Walkden, G., & Breitbarth, A. (2019). Complexity as L2-difficulty: implications for syntactic change. *Theoretical Linguistics*, *45*(3–4), 183–209.

Weerman, F. (1993). The diachronic consequences of first and second language acquisition: the change from OV to VO. *Linguistics*, *31*, 903–931.

Yang, C. D. (2000). Internal and external forces in language change. *Language Variation and Change*, *12*, 231–250.

Yang, C. D. (2002). *Knowledge and learning in natural language.* Oxford: Oxford University Press.

# POTS VERSUS CLOCKS: INTEGRATING ARCHAEOLOGICAL EVIDENCE WITH BAYESIAN ESTIMATES OF DIVERGENCE TIMES IN OCEANIC LANGUAGES

BENEDICT KING*[1], MARY WALWORTH[1], AYMERIC HERMANN[1,2], SIMON J. GREENHILL[1,3] and RUSSELL D. GRAY[1]

*Corresponding Author: benedict_king@eva.mpg.de
[1]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[2]CNRS, Paris, France
[3]Faculty of Science, University of Auckland, New Zealand

Bayesian phylogenetics provides a quantitative method with which to estimate the timing of language divergences (Hoffmann *et al.* 2021). In such analyses, a "relaxed clock" of lexical change is used to date nodes in the language tree. A relaxed clock assumes an approximately predictable rate of lexical innovation, while allowing the rate to vary across the branches of the tree. In order to date a phylogenetic tree, the clock needs to be calibrated by integrating prior information to constrain the age of particular nodes in the tree (Maurits *et al.* 2020). The expansion of Austronesian speakers across Oceania is a topic of research in which Bayesian methods should provide valuable insight. Since the publication of the Austronesian phylogenetic tree of Gray *et al.* (2009), the Austronesian Basic Vocabulary Database (Greenhill *et al.* 2008) has undergone significant expansion. New data, combined with improvements in analysis methods and new archaeological dates should provide further insights into the timing and expansion of Austronesians into Oceania.

To obtain a time-scaled tree of Oceanic languages, we analysed lexical data in the form of cognate sets in the Bayesian phylogenetic software BEAST2 (Bouckaert *et al.* 2018). The tree was time-scaled using a relaxed clock and calibration distributions on the divergence of the Oceanic, Polynesian and Eastern Polynesian languages. Despite the relatively informative prior probability distributions on divergence time, our analysis produced an estimate for the divergence of Oceanic languages that strongly conflicted with the archaeological record (Fig. 1), a result that holds for analyses using a variety of different models of tree shape. This ancient date estimated for Oceanic languages evokes the "Rocks versus Clocks" debate in evolutionary biology, in which molecular clocks frequently produce ancient divergence dates that can predate the earliest fossil record of a group by 10s or 100s of millions of years (Benton 1999, Puttick *et al.* 2016). Surprisingly, these older dates for the Oceanic tree are found despite no apparent signal in the lexical data supporting such a divergence.

We investigated the ability of relaxed clock analyses to recover dates within an age range consistent with the archaeological data using a series of simulations. Data was simulated on a tree taken from an analysis in which the divergence time for Oceanic was enforced to be no more than 3300 years. Although the relaxed clock performed reasonably when simulations were performed according to randomised branch rates, we found that when data were simulated using the observed branch rates, the age of Oceanic was consistently overestimated, as were other nodes in the phylogeny. The clock model showed some degree of statistical inconsistency, in that larger simulated datasets returned more inaccurate age estimates. We also found that the relaxed clock was unable to correctly estimate the rates of lexical innovation on branches, consistently underestimating rates on branches simulated with a rapid rate. The relatively old date estimated for Oceanic may therefore be a statistical artefact resulting from a high rate of lexical innovation at the origin of the Central Pacific subgroup of Oceanic.

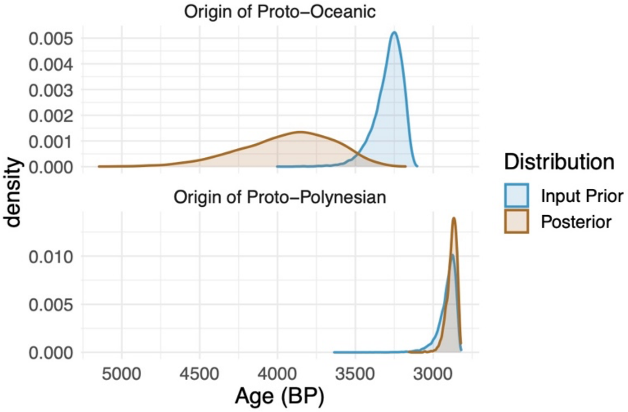


Figure 1: Comparison of prior age calibrations (based on archaeological data) with ages estimated from the data (posterior). For Oceanic as a whole, the estimated age strongly conflicts with the prior.

The inability of these analyses to recover the known age of Oceanic in a simulation scenario calls into question the accuracy of ancient dates estimated from relaxed clocks both in linguistics and evolutionary biology. In both fields, groups chosen for study might often be expected to show rapid rates of evolution (lexical or molecular) at the base of the tree (Beaulieu, 2016), for example due to migration to new territories or recovery from mass extinctions. Our study suggests that relaxed clocks overestimate the age of linguistic groups in at least some situations.

## References

Beaulieu, J. M., O'Meara, B. C., Crane, P., & Donoghue, M. J. (2015). Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Systematic biology, 64*(5), 869-878.

Benton, M. J. (1999). Early origins of modern birds and mammals: molecules vs. morphology. *BioEssays*, *21*(12), 1043–1051.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., ... & Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *15*(4), e1006650.

Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, *323*(5913), 479–483.

Greenhill, S. J., Blust, R., & Gray, R. D. (2008). The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evolutionary Bioinformatics*, *4*, EBO-S893.

Hoffmann, K., Bouckaert, R., Greenhill, S. J., & Kühnert, D. (2021). Bayesian phylogenetic analysis of linguistic data using BEAST. *Journal of Language Evolution*, *6*(2), 119–135.

Maurits, L., de Heer, M., Honkola, T., Dunn, M., & Vesakoski, O. (2020). Best practices in justifying calibrations for dating language families. *Journal of Language Evolution*, *5*(1), 17–38.

Puttick, M. N., Thomas, G. H., & Benton, M. J. (2016). Dating placentalia: Morphological clocks fail to close the molecular fossil gap. *Evolution*, *70*(4), 873–886.

# COMPOSITIONALITY ARISES FROM ITERATED LEARNING DESPITE A PREFERENCE FOR HOLISTIC SIGNALS: AN EXPERIMENTAL MODEL OF SIGN LANGUAGE EMERGENCE

SIMON KIRBY*[1], YASAMIN MOTAMEDI[1], LUCIE WOLTERS[1], and KENNY SMITH[1]

*Corresponding Author: simon.kirby@ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, UK

**Background** Cultural evolution has been proposed to account for the origins of compositionality (Smith, 2018), by which we mean the use of a segmented sequence of signal elements, each of which encodes part of the meaning of the whole signal. Often in evolutionary models and experiments, the starting state of a language is assumed to be holistic, lacking sequential segmentation of signals into meaningful sub-parts. Ultimately, a cognitive bias is assumed, which leads to the replacement of a holistic starting state with a compositional one. But if a bias against holistic representations exists, why would they ever arise in the first place?

We use the example of the gradual emergence of segmented manner/path representations in Nicaraguan Sign Language (Senghas, Kita, & Ozyürek, 2004) to investigate whether this kind of segmented compositionality can arise through iterated learning *even if there is an initial preference in favour of non-segmented representations*. We use a series of large-scale experiments (N=1700) in which participants choose between two gestures describing events, specifically videos of balls moving along different paths and in different manners, (Schouwstra, Abramova, Motamedi, Smith, & Kirby, 2014). The gestures either describe the video by encoding manner and path simultaneously, or by segmenting them (e.g. gesturing the path, then the manner).

**Experiments** In **experiment 1**, participants (N=100) were presented with 16 videos of balls moving with four different paths (e.g., down a slope, around a circle) and four different manners (e.g., bouncing, sliding). Each video was accompanied by two gesture videos representing the movement iconically, with manner and path either segmented or simultaneous. For each of the 16 videos, participants chose which gesture video best conveyed the event. The results show a strong preference for simultaneous gesture ($\beta = 2.30, SE = 0.34, z = 6.83, p < 0.001$), replicating findings from previous silent gesture improvisation experiments (Clay, Pople, Hood, & Kita, 2014), and co-speech gesture (Senghas et al., 2004).

Next we tested whether this clear preference for simultaneous gesture is replicated in learning. In **experiment 2**, a new set of participants (N=100) went through the same procedure as before except that they were first exposed to 12 training trials in which a sample of the ball videos were presented alongside a gesture video. Half of the training trials appeared with segmented gestures, half with simultaneous ones. In contrast to the preferences without training, we now see a preference for *segmented* gestures ($\beta = -0.82, SE = 0.28, z = -2.97, p = 0.003$), and a significant difference between experiment 1 and 2, ($\beta = -2.83, SE = 0.29, z = -9.60, p < 0.001$).

Although this difference between preferences with and without exposure is striking, the preference for segmented gestures after learning is not as strong as the preference for simultaneous gestures without learning. In NSL, the use of segmentation increases over cohorts, suggesting accumulation through cultural evolution. In **experiment 3**, we ran 100 iterated learning chains using the learning paradigm from experiment 2 in which the participants' choices at generation $g$ is the training data for participants at generation $g + 1$, using the data from experiment 2 as the first generation and continuing for a further 7 generations (N=700). We find that the preference for segmented gestures is amplified over generations ($\beta = -0.10, SE = 0.05, z = -2.36, p = 0.02$).

These results are compelling evidence that iterated learning can lead to languages that reflect biases that are the *opposite* of preferences of participants prior to learning. However, to fully model the sign language emergence context, where segmented forms emerge following an initial preference for holistic structures, in **experiment 4** (N=800) we replicated experiment 3 using the output of the baseline preference from experiment 1 as seed rather than the 50:50 exposure of experiment 2. Here too we see a significant accumulation of segmented gestures over generations ($\beta = -0.17, SE = 0.05, z = -3.68, p < 0.001$).

**Discussion** We can treat language evolution as a Markov process and use all the data from experiments 2, 3 and 4 to estimate the full transition matrix from language to language. From this we derive the stationary distribution for languages. This shows a preponderance of segmented languages, but with some probability mass on simultaneous ones, suggesting that although cultural evolution will tend to lead languages towards segmented compositionality, some variability can nevertheless be expected across sign languages.

These results demonstrate that cultural evolution through iterated learning can lead to outcomes that appear to run counter to preferences prior to iterated learning. We use this to introduce a crucial distinction between *naturalness biases* which affect signals that are not part of a set of conventions, and *systematicity biases* that arise when signals are learned as part of a larger set.

## References

Clay, Z., Pople, S., Hood, B., & Kita, S. (2014). Young children make their gestural communication systems more language-like: Segmentation and linearization of semantic elements in motion events. *Psychological science*, *25*(8), 1518–1525.

Schouwstra, M., Abramova, K., Motamedi, Y., Smith, K., & Kirby, S. (2014). From silent gesture to artificial sign languages. In *Workshop on evolution of signals, speech and signs, Evolang X* (p. 22).

Senghas, A., Kita, S., & Ozyürek, A. (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science (New York, N.Y.)*, *305*(5691), 1779–82.

Smith, K. (2018). The cognitive prerequisites for language: insights from iterated learning. *Current opinion in behavioral sciences*, *21*, 154–160.

# THE RECURSION
# THAT THOSE WHO STUDY LINGUISTICS SEEK
# IS EVERYWHERE IF YOU KNOW WHERE TO LOOK FOR IT

Robert Kluender*1, Emily Davis*1, and Marc Rolnik

*Corresponding Author: e4davis@ucsd.edu
1Department of Linguistics, UC San Diego, La Jolla, CA, USA

In this paper we build on previous claims that storytelling, music, and dance/performative gesture comprise intrinsic parts of human ritual, occurring ubiquitously across cultures and across time (Staal, 1979, 1980, 1984; Merker, 1999, 2009; Lewis, 2018). Ritual is unique in its emphasis on structural form and proper execution over semantic/functional content. Staal claimed that Vedic ritual in particular was recursive, based on its use of center-embedded symmetries. Here we trace the ubiquity of such structural principles throughout history at every level of analysis in literary (frame narrative, ring composition, chiasma) and musical composition (retrograde, sonata/minuet/rondo, arch form, twelve-tone) to argue that these persist as pervasive remnants of ritual culture. We then present the results of a corpus study showing that center embedding is more frequent in head-final languages, and raise the possibility that even in head-initial languages, center embedding may be more common than assumed. Taken together, the evidence we present suggests that center embedding is not only alive and well in human language and culture today, but that it has also likely been prevalent throughout the history of our species.

## 1. Introduction

Twenty years ago, Hauser, Chomsky & Fitch (2002) threw down the gauntlet for the study of language evolution with the strong claim that recursion (Fitch & Martins 2014:98, “the same hierarchical structure is repeated at multiple levels of the hierarchy”) is the core property and *sine qua non* of the language faculty, and that it cannot have been an adaptive property that was selected for (2002:1572-3). Yet the concession is also made that recursion in language could have been exapted from other, non-communicative domains of animal cognition, such as navigation, number quantification, or social relationships (2002:1571,1578).

Here we take a slightly different tack in the quest for recursion by expanding on proposals by Staal (1979, 1980, 1984) highlighting the similarity of recursive properties in Vedic ritual and in language. Of particular note is Staal’s claim that the center embedding evident in ritual may have served as a template for recursion in human language—and not the other way around. In partial support of this idea, animals as cladistically diverse as Bengalese finches (Abe & Watanabe, 2011) and macaque monkeys (Jiang et al., 2018) have been trained to recognize recursive, center-embedded patterns; language therefore cannot be a necessary condition for this ability. Yet if the center embedding found in ritual played such a central role in the evolution of linguistic recursion, (1) **why would center embedding appear to be so infrequent today** (Karlsson, 2007; Levinson, 2013), and (2) **shouldn't there likewise be more traceable remnants of center embedding in human culture**? These are the questions we address in this paper. and we present what we find to be striking evidence along these lines.

With regard to (1), Staal (1980) proposed that the much more luxurious time scale on which ritual typically unfolds accommodates the complexity of center embedding without difficulty, whereas the speeded pace of spoken or signed language communication in real time gives rise to obvious problems of verbal working memory, as first pointed out by Chomsky & Miller (1963; see also Warren & Gibson, 2002, regarding reference form). However, given the apparent ease with which native speakers process head-final languages (e.g. Inoue & Fodor, 1995), we predicted that head-final languages should exhibit more center-embedding than head-initial or mixed-headedness languages, and conducted a corpus search to this end. We also assembled a collection of center-embedded examples from popular media in English. These results are found in section 4.

With regard to question (2) above, Lewis (2018) has described hunter-gatherer ritual as a communicative, transactional relationship between a society and the forces of nature, manifesting in institutionalized communal reenactments consisting of storytelling, music, and dance (with specified melodies, songs, and rhythms). This is remarkably consistent with traditional descriptions by Sanskrit scholars of the more formalized, heavily constrained, and institutionally codified ritual practice of Vedic culture. This in turn suggests that these may well be universal features of human ritual. Merker (1999, 2009) draws similar conclusions: "Presumably, the vocal, the gestural and the social aspects of group display were never separate, nor are they kept separate in the conceptual categories of a number of non-Western peoples even today. Their languages subsume rhythm, song, dance and ritual (celebration) under a unitary concept…." (Merker, 2009:53; see also Cassirer, 1946:40-41). More crucially for our purposes, Lewis (2018) emphasizes that the fundamental backbone and crucial focus of hunter-gatherer ritual is not its semantic content, but its specified acoustic properties, structural form, and manner of execution. This is precisely Staal's (1979, 1980, 1984) claim with regard to the structure of Vedic ritual as well.

We build on such cross-cultural similarities to show that center embedding persists and is well attested in the obvious descendants of early human ritual culture today: structural features of (especially Western[1]) literary (section 2) and music composition (section 3). Taken together, the evidence we present suggests that not only is center embedding alive and well in human culture and cognition today, but that it has likely been prevalent throughout human history—and arguably in some prior hominin species as well. As such, the ubiquity of center embedding in human culture offers a tractable avenue for recursion to have gained a foothold in language via its common association with ritual and music.

## 2. Center embedding in epic poetry

It turns out that not only is Vedic ritual center-embedded: the oral literature with which it is intertwined is as well. The so-called "frame story" structure of Indian epics like the *Mahābhārata*, in which stories are repeatedly embedded one within one the other, has been shown to reflect the structure of the rituals that serve as the narrative framework for the recounting of those tales (Minkowski, 1989). Alternatively, the epic embedding of one myth within another may be designed to elucidate and justify the particulars of ritual practice (Witzel, 1987). However, ABC…X…CBA narrative structure is not limited to Vedic culture, but ubiquitous across Indo-European epic poetry (Watkins, 1995). Classicists refer to it as ring composition (van Otterlo, 1944). This symmetrical structure has been identified in the *Iliad* and the *Odyssey* (van Otterlo, 1944; Whitman, 1958; Haig Glaser, 1969), the *Aeneid* (Duckworth, 1962), *La Chanson de Roland* (Niles, 1973), *Beowulf* (Niles, 1979), and in books of the Hebrew Bible (Fishbane, 1975; Alter, 1987; Rosenberg 1987; Douglas 2007). The same structuring device has been claimed for literary works as diverse as *1001/The Arabian Nights* (whose core tales are recognized to be of Indian and Persian origins; Irwin, 1994), Boccaccio's *Decameron*, Chaucer's *The Canterbury Tales*, Shakespeare's plays (Rose, 1972), Mary Shelley's *Frankenstein*, and Sterne's *Tristram Shandy* (Douglas, 2007).[2]

Importantly, however, the literary technique of center embedding applies not only at the level of entire works, but also to the more fine-grained structure of individual hymns, poems, or verses. It is instantiated and signaled by a number of both narrative and linguistic bracketing devices, including case, tense/aspect, person, mood, connectives, and specific lexical forms (Minkowski, 1989). For just one such example, so-called riddle hymns of the *Ṛg Veda* have a recognizable structure in which the enigma itself is embedded in a verse in the middle of the hymn, with parallel and symmetric stylistic and lexical bracketing devices on

---

[1] For literature at least, Douglas (2007:5,8) cites further examples from Madagascar, China, Burma (Myanmar), Thailand, Vietnam, Malaysia, Indonesia, Trans-New Guinea, and Hawai'i.

[2] See Dane (1993) for a general critique of analyses of classical and medieval literature that rely on ring composition as an organizing principle.

either side (Brereton, 1999; Jamison, 2004); Douglas 2007:36-37 includes both of these conventions in her criteria for ring composition). Likewise, at the level of individual verses or lines, chiasmus, the symmetrical (ABBA) structuring of phrases or clauses via lexical and/or syntactic means, is a common poetic device.

## 3. Center embedding in musical composition

The same structural principle is prevalent in the composition of Western music. Palindromic or mirror patterns in music are called retrograde, meaning that the music is identical on some parameter when performed both forward or backward.[3] Such early forms do exist, for example a 14th century rondeau by Machaut and a 16th century canon by Byrd. However, these are exceedingly rare, especially from the Baroque through the Romantic eras, due to the homophony that dominated from the 17th through 19th centuries: chordal structure supporting an independent melody, based on a series of harmonic progressions going from dissonance to consonance. That said, as might be expected, eminent formalists like Bach and Haydn both toyed with strictly palindromic composition in individual works.

However, musicologists have since expanded the definition of hierarchical structure in music to include composition at every level of analysis, from multi-movement works, individual movements, and passages within movements to musical gestures such as themes, motifs, rhythm, texture, and color. Recursive structure provides the architectural basis for the sonata (ABA), so-called ternary forms (ABA) like the minuet, and the rondo (ABACABA). Recursiveness also manifests in the arch form, in which passages or movements are mirrored on the basis of key, tonal center, or contour. A variety of well-known composers from Beethoven and Schumann through Mahler and Bartok composed in arch form in some of their works. Rohrmeier (2011) has analyzed tonal harmonic progressions as compatible with a context-free grammar. Palindromic permutations were also fundamental to the 2nd (twelve-tone) Viennese school, the later return of tonality in minimalism (Porter 1971), and 21st century eclecticism. Symmetry has thus played a central role in the history of Western music as well as in its literature.

## 4. How frequent is center embedding?

The received wisdom in linguistics and psycholinguistics is that center embedding is uncommon because of the burden it imposes on working memory resources

---

[3] Interestingly, this is one (and the most difficult) of the techniques—called *ghana-pāṭha* recitation—employed by Vedic pandits to learn by oral means the vast repertoire of hymns they are required to commit to memory. Individual words of the hymn in various combinations are repeated forward and backward, with total disregard for syntactic or semantic well-formedness (Egenes 1989:48).

(Chomsky & Miller, 1963), as also noted by Staal (1980).[4] Yet given the apparent ubiquity of center embedding in other cultural domains derived from ritual, we hypothesized that center embedding may actually be more common in language than is generally assumed. For instance, center embedding is both more tolerable and more common in writing than in speech, thanks to the use of externalized cognitive representations that are stable over time (Karlsson, 2007). This impression is buttressed by the number of naturally occurring examples we encounter in our casual reading and in corpora. Here is just one such example:

> …the idea [that only the people
> [who heard the tale straight from Homer's lips]
> had the authentic experience of the epics] is facially absurd.

https://forums.somethingawful.com/showthread.php?threadid=3884594&userid=0&perpage=40&pagenumber=5#post493476903

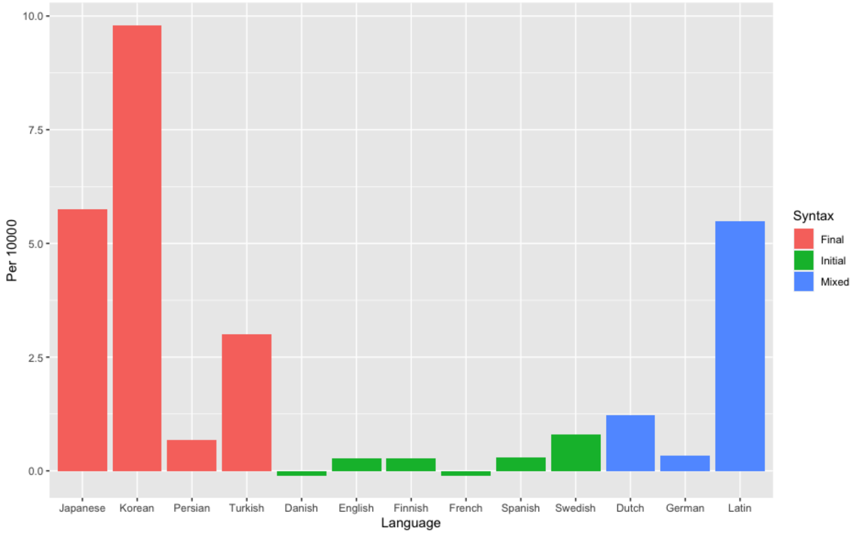


Figure 1. The occurrence per 10,000 sentences of center embedded structures in corpus searches of head-final and head-initial languages, and of languages with mixed headedness.

However, the bulk of the linguistic literature on center embedding focuses on English and other European languages (e.g. Karlsson, 2007). We predicted that center embedding should be more common in head-final languages, given that clausal complementation with SOV word order favors center-embedded structure, and subjects are therefore more frequently separated from their verbs anyway. We

---

[4] Levinson (2013) has claimed that center embedding is more common in conversational discourse, but see Legate et al. (2014) for a critique of Levinson’s claims regarding syntactic recursion.

searched the UD database (Nivre et al., 2018) for center-embedded structures in a variety of head-final and head-initial languages, as well as in languages with mixed headedness. With the help of native speaker linguists, we eliminated sentences with incorrect syntactic tagging and parenthetical clauses. Fisher's exact tests confirmed a significant 3-way difference: head-final languages > mixed headedness languages > head-initial languages (Figure 1; all $p < .0001$).

## 5. Conclusion

Starting from the rather uncontroversial notion that storytelling, music, song and dance are inextricably bound up in human ritual as a unified whole (Lewis, 2018; Merker, 2009), we have attempted to show that the purported center-embedded structure of ritual itself (Staal, 1979, 1980) is likewise an intrinsic part of (at least Western) epic poetry and classical music. Our suggestion is that these remnants of ancient ritual reflect a human preoccupation with symmetric structure (Douglas, 2007). Merker (2009:56) goes so far as to characterize language as a meta-ritual: “Language shares with the ritual mode out of which it grew an insistence on proper form, yet differs from it by its emancipation from the finite particularity of rituals.” It has been claimed that center embedding is more prominent both in conversational discourse (Levinson 2013) and in writing (Karlsson 2007). We have shown that it is in fact more common in (the writing of) head-final than in head-initial languages, and we suspect that it is also more common than corpus studies would suggest, even in head-initial languages.

Taken as a whole, our proposal meshes well with others that seek the source of recursion in human language across levels of linguistic analysis, cognitive domains, and spheres of human (and animal) cultural activity. Our findings simultaneously broaden the body of available evidence for such proposals, help to contextualize them in a larger cultural framework, and sharpen the empirical linguistic facts relevant to the use of center embedding in human language.

## References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, *14(8)*, 1067.

Alter, R. (1987). The characteristics of ancient Hebrew poetry. In R. Alter, & F. Kermode (Eds.), *The literary guide to the Bible* (pp. 611-624). Cambridge, MA: Harvard University Press.

Brereton, J.P. (1999). Edifying puzzlement: Ṛgveda 10.129 and the uses of enigma. *Journal of the American Oriental Society*, *119(2)*, 248-260.

Cassirer, E. (1946). *Language and myth.* New York, NY: Dover Publications.

Chomsky, N., & Miller, G.A. (1963). Introduction to the formal analysis of natural languages. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology, 2* (pp. 269-321). New York, NY: Wiley & Sons.

Dane, J.A. (1993). The notion of ring composition in classical and medieval studies: A comment on critical method and illusion. *Neuphilologische Mitteilungen*, *94(1)*, 61-67.

Douglas, M. (2007). *Thinking in circles: An essay on ring composition.* New Haven, CT: Yale University Press.

Duckworth, G.E. (1962). *Structural patterns and proportions in Vergil's Aeneid: A study in mathematical composition.* Ann Arbor, MI: University of Michigan Press.

Egenes, T. (1989). *Introduction to Sanskrit: Part Two*. Delhi: Motilal Banarsidass Publishers.

Fishbane, M. (1975). Composition and structure in the Jacob cycle (Genesis 25: 19-35:22). *Journal of Jewish Studies*, *26(1-2)*, 15-38.

Fitch, W.T., & Martins, M.D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316(1)*, 87-104.

Haig Glaser, J. (1969). A structural analysis of the digressions in the *Iliad* and the *Odyssey*. *Harvard Studies in Classical Philology*, *73*, 1-43.

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? S*cience*, *298*, 1569-1579.

Inoue, A., & Fodor, J.D. (1995). Information-paced parsing of Japanese. In R. Mazuka, & N. Nagai (Eds.), *Japanese sentence processing* (pp. 9-63).

Irwin, R. (1994). *The Arabian nights: A companion.* London: Penguin Books.

Jamison, S.W. (2004). Poetry and purpose in the Ṛgveda: Structuring enigmas. In A. Griffiths, & J.E.M. Houben (Eds.), *The Vedas: Texts, language and ritual. Proceedings of the 3rd International Vedic Workshop* (pp. 237-249). Groningen: Egbert Forsten.

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., & Wang, L. (2018). Production of supra-regular spatial sequences by macaque monkeys. *Current Biology, 28(12)*, 1851-1859.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*, 365-392.

Legate, J.A., Pesetsky, D., & Yang, C. (2014). Recursive misrepresentations: A reply to Levinson (2013). *Language*, *90(2)*, 515-528.

Levinson, S.C. (2013). Recursion in pragmatics. *Language*, *89*, 149-159.

Lewis, J. (2018). Music before language: Observations from a hunter-gatherer's point of view. Plenary address, EvoLang XII, Toruń, Poland, April 19. *https://www.youtube.com/watch?v=xkwY84mfhNg&t=1889s*

Merker, B. (1999). Synchronous chorusing and the origins of music. *Musicae*

*Scientiae*, *3*(1_suppl), 59-73.

Merker, B. (2009). Ritual foundations of human uniqueness. In S. Malloch, & C. Trevathen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 45-59). Oxford: Oxford University Press.

Minkowski, C.Z. (1989). Janamejaya's *Sattr*a and ritual structure. *Journal of the American Oriental Society*, *109(3)*, 401-420.

Niles, J.D. (1973). Ring-composition in *Le Chanson de Roland* and *Le Chançun de Willame*. *Olifant*, *1(2)*, 4-12.

Niles, J.D. (1979). Ring composition and the structure of Beowulf. *PLMA 94(5)*, 924-935.

Porter, D.H. (1971). Ring-composition in classical literature and contemporary music. *The Classical World*, *65(1)*, 1-3, 6, 8.

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5(1)*, 35-53.

Rose, M. (1972). *Shakespearean design*. Cambridge, MA: Harvard University Press.

Rosenberg, J. (1987). Jeremiah and Ezekiel. In R. Alter, & F. Kermode (Eds.), *The literary guide to the Bible* (pp. 184-206). Cambridge, MA: Harvard University Press.

Staal, F. (1979). The meaninglessness of ritual. *Numen*, *26(1)*, 2-22.

Staal, F. (1980). Ritual syntax. In M. Nagatoni, B.K. Matilal, J.M. Masson, & E.C. Dimock, Jr. (Eds.), *Sanskrit and Indian studies* (pp. 119-142). Dordrecht: D. Reidel Publishing Company.

Staal, F. (1984). Ritual, mantras and the origin of language. In S. D. Joshi (Ed.), *Amṛtadhārā: Professor R.N. Dandekar Felicitation Volume* (pp. 403-425). Delhi: Ajanta Publications.

Van Otterlo, W.A.A. (1944). Untersuchungen über Begriff, Anwendung, und Entstehung der griechischen Ringkomposition. *Medeleelingen der Koninklijke Nederlandsche Akademie van Wetenschappen (1-6)*, 131-176.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*, 79-112.

Watkins, C. (1995). *How to kill a dragon: Aspects of Indo-European poetics*. Oxford: Oxford University Press.

Whitman, C.H. (1958). *Homer and the heroic tradition.* Cambridge, MA: Harvard University Press.

Witzel, M. (1987). On the origin of the literary device of the "frame story" in Old Indian literature. In H. Falk (Ed.), *Hinduismus und Buddhismus: Festschrift für Ulrich Schneider* (pp. 380-414). Freiburg: Hedwig Falk Verlag.

# ECOLOGICAL AND SOCIO-CULTURAL CONSTRAINTS ON LINGUISTIC DIVERSIFICATION: NEW INSIGHTS FROM THE BANTU EXPANSION

EZEQUIEL KOILE [*1], SIMON J GREENHILL [1,2], DAMIAN E BLASI [3, 1,4],
REMCO BOUCKAERT [5] and RUSSELL D GRAY [1,6]

[*]Corresponding Author: ezequiel_koile@eva.mpg.de
[1]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[2] School of Biological Sciences, University of Auckland, Auckland, New Zealand
[3] Department of Human Evolutionary Biology, Harvard University, Peabody Museum, Cambridge, MA, USA
[4] Human Relations Area Files, Yale University, New Haven, CT, USA
[5] Centre for Computational Evolution, University of Auckland, New Zealand
[6] School of Psychology, University of Auckland, Auckland, New Zealand

The Bantu expansion was a massive migration that reshaped the linguistic, economic, and cultural landscape of Africa. It led to the proliferation of Bantu-speaking populations throughout sub-Saharan Africa and today more than 500 Bantu languages are spoken by 240 million people across an area of 9 million square kilometers (de Filippo et al. 2012). This expansion has been associated with major economic and cultural changes across sub-Saharan Africa, including a more sedentary way of life, iron working, and crop cultivation (Neumann et al. 2012, Grollemund et al. 2015, Currie et al. 2013).

Although there is a consensus about the time and location of the homeland of the Bantu people, around 5,000 years before present, near the border between current Nigeria and Cameroon, by the Guinea Gulf (Diamond et al. 2003, Blench 2006, Vansina 1995), substantial uncertainty remains about the route and environmental conditions faced by early Bantu speakers during their migrations. Until recently, it was believed that these populations, characterized by their agriculturalist subsistence method, were unable to adapt to the West African Rainforest, which is located in the way of their migration paths, according to certain hypotheses.

We use the recently developed “break-away” geographical diffusion model (Bouckaert et al 2018), specially designed for modeling migrations, together with "augmented" geographic information, in order to reconstruct the Bantu language family expansion. This Bayesian phylogeographic approach (Bouckaert et al. 2014) with augmented geographical data provides a powerful way of linking linguistic, archeological and genetic data to test hypotheses about large language family expansions.

We find that our analyses support the hypothesis of an expansion through central African tropical forests at 4,420 BP (4,040-5,000 95% HPDI), well before the savanna corridor known as the Sangha River Interval was open in its interior. This is consistent with a slow adaptation of the Bantu speaking populations to the rainforest, where interaction with their hunter-gatherer neighbors was fundamental (Klieman 2003).

We take these results to show a wider trend in the study of language diversification: While subsistence has shaped the expansion and the tempo of many language families, it does not impose an inescapable barrier to alternative diversification patterns. This is critical for understanding the role of human flexibility and ingenuity when thinking broadly about the processes underlying language evolution.

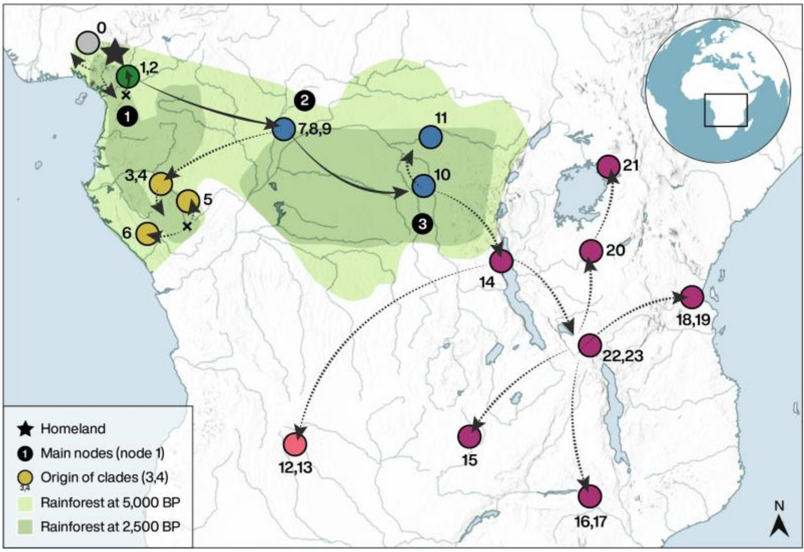


Figure 1. Bantu migrations reconstructed by using the break-away model in the augmented phylogeographic tree in Figure 1. The homeland is marked with a star, and main nodes are numbered (1-3), as well as main clades (0-23). Each circle represents the median value of the posterior distribution for the origin of the respective clade, and their colors represent the geographical region spanned by the corresponding languages nowadays. The span of the rainforest at 5,000 BP and at 2,500 BP is shown, according to (Maley 2001, Maley 2002).

## References

Blench, Roger. 2006. Archaeology, Language and the African Past. Lanham: Altamira Press.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard, MA., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A Software Platform

for Bayesian Evolutionary Analysis. PLoS Computational Biology, 10(4), e1003537. doi:10.1371/journal.pcbi.1003537

Bouckaert R., Bowern C., Atkinson Q.
The origin and expansion of Pama-Nyungan languages across Australia
Nature Ecology & Evolution, 2018

Currie TE, Meade A, Guillon M, Mace R. Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. Proceedings of the Royal Society B: Biological Sciences. 2013;280(1762):20130695. doi:10.1098/rspb.2013.0695.

de Filippo, Cesare, Koen Bostoen, Mark Stoneking, Brigitte Pakendorf,
Bringing together linguistic and genetic evidence to test the Bantu expansion
Proc. R. Soc. B 2012 279 3256-3263; DOI: 10.1098/rspb.2012.0318.
Published 11 July 2012

Diamond J., Bellwood P. 2003 Farmers and their languages: the first expansions. Science 300, 597–603. doi:10.1126/science.1078208 (doi:10.1126/science.1078208)

Grollemund, R., Branford, B., Bostoen, K., Meade, A., Venditti, C. & Pagel, M., 2015. "Bantu expansion shows habitat alters the route and pace of human dispersals". Proceedings of the National Academy of Sciences (PNAS), 112:43, pp. 13296-13301.

Klieman, K. A. (2003). ``The pygmies were our compass": Bantu and Batwa in the history of west central Africa, early times to c. 1900 CE. Greenwood.

Maley J (2001) La destruction catastrophique des forêts d’Afrique centrale survenue il ya environ 2500 ans exerce encore une influence majeure sur la répartition actuelle des formations végétales. Syst Geogr Plants 71:777–796

Maley J (2002) A catastrophic destruction of African forests about 2,500 years ago still exerts a major influence on present vegetation formations. IDS Bull 33(1):13–30.

Neumann, K., Bostoen, K., Höhn, A., Kahlheber, S., Ngomanda, A., & Tchiengué, B. (2012). First farmers in the Central African rainforest: A view from southern Cameroon. Quaternary International, 249, 53-62.

Vansina, Jan. 1995. New linguistic evidence and the Bantu expansion. Journal of African History 36: 173–195.

# THE EVOLUTIONARY ROOTS OF HUMAN COMMUNICATIVE REPAIR: THE CASE OF CHIMPANZEE GROOMING

KAYLA KOLFF*[1] and SIMONE PIKA[1]

*Corresponding Author: kkolff@uni-osnabrueck.de
[1]Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

Human communication is characterized by several distinct aspects such as directionality, reference, and role reversal. If a communicative interaction fails, humans use distinct mechanisms to repair it. Surprisingly, relatively little is known about the evolution of repair and possible precursors in non-linguistic species. Thus, this paper introduces possible precursors of communicative repair in grooming interactions of one of our closest living relatives, the chimpanzees (*Pan troglodytes schweinfurthii*). In 1996, Dunbar had suggested that early humans engaged in complex grooming interactions, before grooming was replaced with a more efficient bonding system - language. Therefore, grooming offers a unique platform to investigate the evolutionary precursors of communicative repair.

## 1. Introduction

How humans perceive and interact with each other is predominantly through language. Why only humans developed this unique communicative system remains a mystery (Hauser et al., 2014; Knight, Studdert-Kennedy, & Hurford, 2000) and has led to considerable scientific debate (Christiansen & Kirby, 2003). Some researchers postulated that turn-taking is one of the most ancient mechanisms underpinning the layered language system (Levinson, 2016). This cooperative interaction during conversations was first systematically described by Sacks and colleagues (1974). It requires the combined effort of at least two participants, who alternate short, reciprocal, flexible, and non-overlapping turns to coordinate their interaction during conversations. Turn-taking involves distinct temporal relationships (~200 ms time windows, Stivers et al., 2009), the establishment of participation frameworks (who and when should they talk, move or act next), and the use of repair mechanisms (hereafter communicative repair) to counteract problems during communication.

Recently, comparative researchers started to investigate whether turn-taking is indeed an ancient mechanism that is already present in non-linguistic primate species (e.g., Fröhlich et al., 2016; Pika, Wilkinson, Kendrick, & Vernes, 2018; Ravignani, Verga, & Greenfield, 2019; Rossano, 2018). Individuals alternating the production of signals can be found in several non-linguistic primate species, with the majority of research focusing on the temporal relationships characterizing turn-taking interactions and adjacency pair-like sequences (see Pika et al., 2018 for a recent review). However, relatively little is known about communicative repair in non-linguistic species.

Communicative repair in human conversations is defined as 'fixing' a misunderstanding or communicative problem (Schegloff, Jefferson, & Sacks, 1977), with researchers distinguishing between self-and-other-initiated repair. Self-initiated repair refers to an actor producing communicative means to counteract trouble in a conversation. It entails *repetition,* e.g., repeating the entire or part of the previously spoken turn, or *elaboration,* e.g., rephrasing the previously spoken turn (Kitzinger, 2012). Dingemanse and colleagues (2015) in a cross-linguistic study, showed that across twelve languages, 48% of communicative repair occurrences included individuals partially repeating their initial spoken turn during spontaneous conversations. In contrast, other-initiated repair refers to someone other than the speaker identifying the communicative trouble and using means to counteract it (Dingemanse & Enfield, 2015). Here individuals may locate the source of trouble in the prior communicative turn (e.g., repeating a certain word), use questions words (e.g., “Who”, “What”, “Where?”, “Huh?”) or maintain silence and stare at the speaker, i.e., freeze looks (Dingemanse & Enfield, 2015; Manrique & Enfield, 2015).

In a study of native English speakers, communicative repair occurred at a delay (~700 ms) after the last spoken turn during face-to-face conversations (Kendrick, 2015). Compared to the average temporal gap of 0-500 ms between turns found across several human cultures (Stivers et al., 2019). Furthermore, Lerner and Raymond (2021) recently showed that communicative repair also occurs in non-verbal bodily actions. They reported troubles during exchanges of manual actions, such as premature (mis)recognition of an emerging action, that are repaired through adjustment of actions (e.g., retarding or reversing an action). Communicative repair, therefore, seems to be crucial for different modes of communication and was suggested as an essential element to facilitate the coordination and cooperation between interacting individuals (Levinson, 2006). However, relatively little is known about the evolution of repair and possible precursors that may be found in non-linguistic species.

### *1.1 Brief overview of communicative repair in non-human primates*

To date, relatively little research attention has been focused on communicative trouble and repair in non-human primates. For instance, Haimoff (1988) examined duets songs (organized bouts comprised of three distinct sequences of vocal behavior between pairs: the introductory, interlude, and great calls) of wild and captive gibbon species (*Hylobates* spp.*)*. On about sixty occasions, gibbons were found to counteract 'errors' (production of sounds not being produced normally by an individual at a point in the bout) by abruptly terminating the duet bout and re-initiating a new sequence. Hence, two clear cases were found where Siamang (*Symphalangus syndactylus*) individuals repaired 'troubles' (not producing the expected call at the appropriate time of a sequence) by ultimately initiating the anticipated call and restarting the sequence. Additionally, Heesen and colleagues (2022) presented two cases on captive bonobos (*Pan paniscus*) and wild chimpanzees (*Pan troglodytes*), where individuals may engage in communicative repair during the context of grooming and respectively joint-travel. They described that self-initiated repair could be found beyond modern humans through signal persistence, repetition, and elaboration, e.g., gesture repetition by a chimpanzee mother to initiate joint-travel. Nonetheless, communicative repair across non-human primates in their natural environment still remains unknown.

Hence, this paper aims to pinpoint possible precursors of communicative repair and to propose a systematic framework to enable cross-comparison between primates. We focused on the interactions of one of our closest living relatives, the chimpanzees (*Pan troglodytes schweinfurthii*) in the wild. Chimpanzees are an ideal species to investigate communicative repair due to their rich vocal and gestural repertoires and high collaboration degrees (Goodall, 1986; Mitani, 2009). Similar structures or traits found between chimpanzees and modern humans may either be homologous relating to shared ancestry (Wrangham, 1987) or convergent evolution of the Pan and Hominin ancestors in encountering similar evolutionary pressures (Tooby, DeVore, & Kinzey, 1987). At the end of the last century, Dunbar (1996) suggested that due to a considerable increase in group size, grooming was replaced in modern humans by the more efficient bonding system of language. Communicative interactions involving repair may therefore have characterized grooming interactions before the dawn of human language. Thus, we centered on interactions in the context of grooming, a frequent behavior in chimpanzee societies and has been suggested as a crucial platform for learning and exchanging communicative signals (Pika, 2009, 2014).

## 2. Communicative repair in wild chimpanzees: Employed signals and temporal relationships

Here we introduce six grooming interactions that involve the production of three distinct gestures, the DIRECTED SCRATCH[1], PRESENT BODY PART, and TOUCH (see Fig. 1). These gestures are commonly produced during grooming interactions (either uni-directional - Fig. 1/A, or bi-directional grooming) by chimpanzees of the Ngogo community, Kibale National Park in Uganda, to request, negotiate and maintain grooming (Pika, 2009, 2014). The interactions involve hallmarks of human communicative interactions, including first-order intentionality, flexibility, directionality, reference, role negotiation, and role reversal (Pika, 2014).

Figure 1. (A) Uni-directional grooming interaction of an adult male (left) grooming another adult male (right). (B) An adolescent male SCRATCHING the back of their head (front) after grooming an adult male (back). (C) An adult male PRESENTING their side (front) and the adult male (back) that was previously grooming him. (D) An adult male (right) TOUCHING the rump of the adult male (left) that he was grooming. All illustrated gestures are depicted in red circles.

The DIRECTED SCRATCH is an auditory and visual signal defined as an individual making a relatively loud and exaggerated scratching movement on the part of their body (Pika, 2014). When an actor (individual grooming the other) signals with a DIRECTED SCRATCH during a grooming interaction (see supplementary materials Fig. S1), the recipient (individual being groomed) could either (i) ignore, (ii) respond by repositioning themself (see supplementary materials Fig. S2 - role negotiation), (iii) groom the actor or (iii) groom the actor at the allocated scratch area (see supplementary materials Fig. S1/C - role reversal) (Pika, 2014). A DIRECTED SCRATCH by an actor stops the ongoing interaction (breaks the temporal relationship). It presents a request to the recipient to give a certain response to potentially ‘fix’ the interaction. For instance, role reversal, where the recipient becomes the actor (see supplementary materials Fig. S1). Or negotiation of the actor’s role where the recipient repositions themself to allow the actor to continue their interaction (see supplementary materials Fig. S2). The potential illustrated repair is initiated by the gestural request of an actor while in parallel disturbing the natural flow and temporal relationship of the grooming interaction (possible self-initiated repair). This forces the recipient to respond and

---

[1] Gestures are depicted from here in SMALL CAPITALS.

act, to continue their interaction, and possibly return to the initial situation (see supplementary materials Fig. S1, where the initial actor after the possible repair ends up being the actor again).

PRESENT BODY PART is a gesture that involves an individual offering a body area such as their arm, back, or rump at a recipient (Pika, 2014). PRESENTING BODY PART is a referential gesture (Hobaiter & Byrne, 2014) because it draws the attention of another individual to a certain location. A PRESENT BODY PART by an actor draws the recipient's attention to request that it is their turn to groom the actor at the presented location. Howbeit, the recipient may still want the actor to continue grooming them and can negotiate this by PRESENTING BODY PART to the actor (see supplementary materials Fig. S3/G). The recipient could also ignore the request by the actor, potentially ending the interaction (see supplementary materials Fig. S4). The recipient negotiating the actor's role may represent a potential other-initiated repair since the recipient communicates that it is still the actor's turn to groom them. Ultimately, the use of PRESENT BODY PART by either an actor or recipient interrupts the natural flow of the grooming interaction.

The TOUCH is a tactile gesture defined as an individual gently putting their hand (<2 seconds) on any body part of the recipient (Pika, 2014) and may accordingly pause (see supplementary materials Fig. S5/B). Simultaneously, the grooming interaction stops, and the involved temporal relationship changes (the flow of grooming interaction is interrupted). This is a directed gesture from an actor to a recipient, where a recipient may respond by orientating themselves, giving access to the body area that the actor touched (i.e., intentional reference request, see supplementary materials Fig. S6/C). The performed request achieves a mutual understanding between the actor and recipient of where to groom next, allowing for the continuation of the interaction. In the scenario where the recipient does not entirely orient themself after a given TOUCH, the actor can orientate themself towards another body area (see supplementary materials Fig. S5/D). Here, the actor does not receive the adequate 'desired' response and counteracts this by performing the wanted request themself (repositioning themself towards a non-groomed body area of the recipient). Similar to the DIRECTED SCRATCH, the TOUCH depicts a possible repair during a grooming interaction. The actor initiates a potential repair, and the recipient can respond by repositioning themself (see supplementary materials Fig. S6). Alternatively, if the recipient's response is inadequate, the actor may pause and reposition themself (see supplementary materials Fig. S5), both cases illustrating possible self-initiated repair.

The gestures DIRECTED SCRATCH, PRESENT BODY PART, and TOUCH produced by chimpanzees changed the temporal relationship of their grooming interaction.

They, therefore, may represent means to 'fix' and eventually prolong the respective interactions, which can be understood as requests "Okay, what now? Here? Your turn or still mine? Didn't I groom you enough". After a given request, the interaction can change towards (1) role reversal (recipient becomes actor, e.g., producing a DIRECTED SCRATCH) or (2) role negotiation (the actor remains the actor, e.g., PRESENTING A BODY PART), or (3) during an actor's turn they can reposition themself following a failed reference of a PRESENTED BODY PART after a given TOUCH. The above scenarios demonstrate how both individuals interact and communicate with each other to continue their grooming interaction and communicate their intentions. In doing so, the actor or recipient can be responsible for employing certain signals to 'fix' and prolong their grooming interaction, which may characterize potential self- and other-initiated repair.

## 3. Conclusion and future trajectory

Despite the growing evidence of distinct elements characterizing human conversational turn-taking also being present in other primates, studies on communicative repair are still minimal. Here, we introduced possible precursors of human communicative repair by focusing on three commonly produced gestures by chimpanzees during grooming interactions to request, negotiate or reverse roles to maintain grooming. Future studies should therefore investigate communicative signals employed by non-human primates during grooming interactions and the outcome of these signals. This will aid in understanding the presence of communicative repair in non-linguistic species and how it relates to modern human communicative repair and the hallmarks of human communication. We, therefore, propose a four-aspect framework (see supplementary materials for further details) within the grooming context: (1) What does the signal entail? (2) Who initiates a signal? (3) What is the outcome after the signal? (4) What is the time window between the signal and outcome?. This framework may offer a systematic approach for future research to compare the evolutionary roots of communicative repair across primates.

## Acknowledgements

We thank the Uganda Wildlife Authority, Uganda National Council of Science and Technology, and Makerere University Biological Field Station for their permission to conduct research. A big thank you to K. Langergraber, J. Mitani, D. Watts and to the Ngogo Field Station staff. We are grateful to the reviewers for their comments. KK was supported by the DFG Research Foundation - project number GRK-2185/1 (DFG Research Training Group Situated Cognition).

## References

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences, 7*(7), 300-307.

Dingemanse, M., & Enfield, N. J. (2015). Other-initiated repair across languages: Towards a typology of conversational structures. *Open Linguistics, 1*, 96-118. doi:10.2478/opli-2014-0007

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., . . . Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PLoS One, 10*(9), e0136100. doi:10.1371/journal.pone.0136100

Dunbar, R. I. M. (1996). *Grooming, Gossip and the Evolution of Language*. London: Faber and Faber Ltd.

Fröhlich, M., Kuchenbuch, P., Müller, G., Fruth, B., Furuichi, T., Wittig, R. M., & Pika, S. (2016). Unpeeling the layers of language: Bonobos and chimpanzees engage in cooperative turn-taking sequences. *Scientific Reports, 6,* 25887. doi:10.1038/srep25887

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of Behaviour*. Cambridge, England: Belknap Press of Harvard University Press.

Haimoff, E. H. (1988). The organization of repair in the songs of gibbons. *Semiotica, 68*(1/2), 89-120.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., . . . Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in Psychology, 5*(401), 12. doi:10.3389/fpsyg.2014.00401

Heesen, R., Fröhlich, M., Sievers, C., Woensdregt, M., & Dingemanse, M. (2022). Coordinating social action: A primer for the cross-species investigation of communicative repair. *Philosophical Transactions of the Royal Society B*. doi:10.1098/rstb.2021.0110

Hobaiter, C., & Byrne, Richard W. (2014). The meanings of chimpanzee gestures. *Current Biology, 24*(14), 1596-1600. doi:10.1016/j.cub.2014.05.066

Kendrick, K. H. (2015). The intersection of turn-taking and repair: The timing of other-initiations of repair in conversation. *Frontiers in Psychology, 6*(250.), 146-161. doi:10.3389/fpsyg.2015.00250

Kitzinger, C. (2012). Repair. In *The Handbook of Conversation Analysis* (pp. 229-256).

Knight, C., Studdert-Kennedy, M., & Hurford, J. R. (2000). *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge: Cambridge University Press.

Lerner, G. H., & Raymond, G. (2021). Body trouble: Some sources of difficulty in the progressive realization of manual action. *Research on Language and Social Interaction, 54*(3), 277-298.

Levinson, S. C. (2006). On the human “interaction engine”. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39-69). Oxford: Berg.

Levinson, S. C. (2016). Turn-taking in human communication – Origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6-14. doi:10.1016/j.tics.2015.10.010

Manrique, E., & Enfield, N. J. (2015). Suspending the next turn as a form of repair initiation: evidence from Argentine Sign Language. *Frontiers in Psychology, 6*(1326), 1-21.

Mitani, J. C. (2009). Cooperation and competition in chimpanzees: Current understanding and future challenges. *Evolutionary Anthropology, 18*(5), 215-227.

Pika, S. (2009). Our grooming cousins: Providing the link to declarative signalling? *Studies in Communication Sciences, 9*(1), 73-102.

Pika, S. (2014). Chimpanzee grooming gestures and sounds: What might they tell us about how language evolved?. In D. Dor, C. Knight, & J. Lewis (Eds.), *The Social Origins of Language: Early Society, Communication and Polymodality* (pp. 129-140). Oxford: Oxford University Press.

Pika, S., Wilkinson, R., Kendrick, K. H., & Vernes, S. C. (2018). Taking turns: Bridging the gap between human and animal communication. *Proceedings of the Royal Society B Biological Sciences, 285*(20180598), 9. doi:10.1098/rspb.2018.0598

Ravignani, A., Verga, L., & Greenfield, M. D. (2019). Interactive rhythms across species: the evolutionary biology of animal chorusing and turn-taking. *Annals of the New York Academy of Sciences, 1453*(1), 12-21.

Rossano, F. (2018). Social manipulation, turn-taking and cooperation in apes. Implications for the evolution of language-based interaction in humans. *Interaction Studies, 19*(1-2), 151-166. doi:10.1075/is.17043.ros

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*(4), 696-735.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*(2), 361-382.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 106*(26), 10587-10592. doi: 10.1073/pnas.0903616106

Tooby, J., DeVore, I., & Kinzey, W. G. (1987). The evolution of human behavior: Primate models.

Wrangham, R. W. (1987). The significance of African apes for reconstructing human social evolution. *The Evolution of Human Behavior: Primate Models.*, 51-71.

# MECHANISMS UNDERLYING THE HERMENEUTIC CIRCLE BETWEEN CONNOTATION AND DENOTATION IN THE CONSTRUCTION OF SYMBOLIC COMMUNICATION SYSTEMS

TAKESHI KONNO*[1], RYUSEI OTA[1], and RYOSUKE NAKANO[1]

*Corresponding Author: konno-tks@neptune.kanazawa-it.ac.jp
[1]Engineering Department, Kanazawa Institute of Technology, Ishikawa, Japan

The expression ”Can you pass the salt?” exchanged at the dining table can be interpreted as either requesting a salt shaker or literally enquiring if the person to whom the question is addressed to is able to pass the salt shaker. An expression has not only a literal meaning (denotation), but also an implied meaning (connotation), and they are interdependent. The process of determining the meaning of an expression involves a saying/implicating circle (Carston, 2002), which is thought to develop eventually into a structure of individual parts and whole, called a hermeneutic circle (Tsuda, 1984; Levinson, 2000). However, the mechanism that lead to the formation of this structure from its primitive stage is unclear. We therefore developed a coordination game with messages that simplifies the denotation and connotation based on the framework of experimental semiotics (Galantucci, 2009; Scott-Phillips & Kirby, 2010). Furthermore, while developing a computational model that is positioned as model-based reinforcement learning that performs as well as humans, we found several mechanisms underlying the primitive circle.

The proposed coordination game is as a terminal-based computer game (Galantucci, 2005) in which players and their partners randomly place pieces in separate $2 \times 2$ rooms and aim to move them to the same room. Before moving a piece, players can choose one of the four available shapes and exchange it with their partners only once at any given time. Players can move their own piece only once or keep it in the same position, but cannot move it diagonally between rooms. The result is then disclosed, and the two pieces are again randomly placed. In the process of repeating this series of operations in each game round, most players form symbol systems that map the shapes to their pieces to the room locations.

In this game, as the pieces cannot be moved diagonally, there is only a $50\%$ chance of matching the partner's move if a piece is moved to a particular room without exchanging messages with the partner. However, if the first sender can tell the location of the present room and the second sender can reply with a destination where the two can meet, they can match their movements. Here, the relationship between game rooms and symbols corresponds to the denotation of a semantic element and the first location and destination of a piece correspond to its connotation of a pragmatic element. In an experiment with 20 pairs of human participants, 14 pairs (65%) constructed a symbol system that could consistently match the rooms to which the pieces were moved by the pair after 60 rounds of repetition. The designed computational model of the proposed game probabilistically determines the symbols, rooms, and when to send symbols. The probability changes through trial-and-error reinforcement learning, and the computational model has a parameter for interpreting symbols as indicating being/going to a room. This parameter is estimated from the partner's action history. In computer simulations using this parameter, the number of rounds required to achieve consistent matching of rooms to which the pieces were moved by both players was reduced to approximately 1/5 of that without using this parameter.

This reduction can be achieved by estimating the location of the partner's piece based on the received symbols and changing the probability of moving the piece to that estimated location. For example, when the location of the partner's piece is known, the probability of moving the piece to a position with no contact should become zero. This mechanism deviates from the framework of reinforcement learning, which changes the behavioral probability based on the rewards obtained from the results. However, this may be the mechanism by which symbols manipulate a partner's behavior. It was also necessary to coordinate the symbols when two players used the same symbol for different rooms or different symbols for the same room. We considered the implicit arrangement of role divisions, such as leader and follower, or the introduction of ambiguity into the denotation to solve this coordination. We adopted the latter mechanism owing to its lower computational cost. The computational model with the above mechanisms could consistently match the rooms to which the pieces were moved by both players for 9 pairs of participants (45%) out of the 20 pairs.

Computational models that deal with both literal and implied meanings can form symbol systems faster than computational models that only deal with literal meanings. In addition, a mechanism that changes the action probability before acting based on the literal and implied meanings inferred from the symbols is necessary for the realization and mutual action coordination at a low cost. These features improve the adaptive value of language acquisition in human evolutionary processes (Scott-Phillips, 2015).

## Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP17H06383.

## References

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication.* John Wiley & Sons.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737–767.

Galantucci, B. (2009). Experimental semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science*, *1*(2), 393–410.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* Cambridge.

Scott-Phillips, T. C. (2015). *Speaking our minds: Why human communication is different, and how language evolved to make it special.* NY: Palgrave Macmillan.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Topics in Cognitive Science*, *14*(9), 411–417.

Tsuda, I. (1984). A hermeneutic process of the brain. *Progress of Theoretical Physics Supplement*, *79*, 241–259.

# FREQUENCY AND CHARACTER COMPLEXITY IN 27 WRITING SYSTEMS

Alexey Koshevoy*[1], Helena Miton[2,3], and Olivier Morin[1,2]

*Corresponding Author: akoshevoy@ens.fr
[1]Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, UMR 8129
[2]Minds and Traditions Group, Max Planck Institute for the Science of Human History, Jena, Germany
[3]Santa Fe Institute, Santa Fe, USA

We are testing whether Zipf's Law of abbreviation — the idea that more frequent symbols in a code are simpler than less frequent ones — is present at the level of individual characters. In its original and most well-known manifestation, the law of abbreviation is a correlation between the length and frequency of spoken or written words. Zipf's Law of Abbreviation has been shown to hold at the level of words in many languages, but it is unknown whether it holds at the level of characters. Characters' complexity is similar to word length, requiring more cognitive and motor effort from producing and processing the more complex symbols. We built a dataset of character complexity and frequency measures covering 27 different writing systems. Our results suggest that Zipf's Law of Abbreviation holds for every single system in our dataset – the more frequent characters have lower degrees of complexity and vice-versa. This result provides further evidence of optimization mechanisms shaping communication systems.

## 1. Introduction

In his pioneering work, George Kingsley Zipf observed that more frequent words tend to be shorter – a principle known as Zipf's Law of Abbreviation (Zipf, 1949). This law-like relation has since been confirmed on data coming from many different languages (see, for example, Piantadosi, Tily, & Gibson, 2011; Bentz & Ferrer-i-Cancho, 2016). Additionally, Zipf's Law of Abbreviation has been shown to arise in artificial language learning experiments (Kanwal, Smith, Culbertson, & Kirby, 2017) and in communication systems of different species (see Ferrer-i-Cancho & Lusseau, 2009; Semple, Hsu, & Agoramoorthy, 2010, amongst many). This effect is usually explained in terms of minimizing cumulative production cost: speakers intend to reduce their average articulation effort. This results in a reduction in the number of sounds that are pronounced overall.

Since scripts can be thought of as a communication system, which maps written characters to phonemes or syllables (Morin, Kelly, & Winters, 2020), we might expect that the same effect will hold for individual characters. Characters do not have length, unlike words. But visual complexity shares several relevant properties with word length. Complex characters take more effort to write and read, just like long words are more effortful for speakers and hearers. Miton and Morin (2021) suggest that characters in writing systems are under similar pressures as words in spoken languages. Rovenchak and Vydrin (2010) have found a negative correlation between the complexity of characters and their frequency in the Nko writing system (West Africa). Similar results were reported for the Vai writing system in (Rovenchak, Mačutek, & Riley, 2008), and Mandarin Chinese characters (Shu, Chen, Anderson, Wu, & Xuan, 2003). The small number of studies that have tested this hypothesis shows a negative association between the complexity and frequency of characters – consistent with Zipf's Law of Abbreviation. However, no large-scale comparative testing was done in this domain. This study fills this gap by using a dataset that consists of 27 writing systems and computational, automated, and replicable measures to quantify character complexity. This is different from the idiosyncratic methods, primarily based on stroke counts used in previous studies (see Changizi & Shimojo, 2005 for an example of such methodology). Large-scale cross-linguistic corpora and datasets containing data on character complexity such as GraphCom (Chang, Chen, & Perfetti, 2018) or the dataset introduced in (Miton & Morin, 2021) offer a way to conduct cross-linguistic research on Zipf's Law of Abbreviation in writing systems.

We hypothesize that writing systems will follow Zipf's Law of Abbreviation. As most writing systems are largely based on handwritten characters shaped by centuries of reproduction, a minimization of the cumulative production cost is expected. There is evidence, based on high-quality but limited data, that writing systems can become less complex over time (Kelly, Winters, Miton, & Morin, 2021), indicating an overall trend towards simplification. Reduction of the complexity of symbols was also observed in interactive graphical communication experiments (Tamariz & Kirby, 2015; Garrod, Fay, Lee, Oberlander, & MacLeod, 2007) We expect that frequency should negatively correlate with complexity, i.e., more frequent characters should have become simpler visually due to pressures to minimize production cost over the course of their constant reproduction.

## 2. Data

### 2.1. *Dataset description*

The dataset used in this study combines complexity measures from (Miton & Morin, 2021) and frequencies for each character. The complexity measures for every character include perimetric complexity and algorithmic complexity. Perimetric complexity was introduced in (Attneave & Arnoult, 1956), and it is defined as the inked surface divided by the perimeter of this inked surface. Miton and Morin (2021) computed this complexity measure using an implementation proposed in (Watson, 2012). Additionally, Pelli, Burns, Farell, and Moore-Page (2006) demonstrate that perimetric complexity closely correlates with human visual processing effort. Algorithmic complexity is the number of bytes needed to store a compressed version of the character. The frequencies of individual characters were obtained from biblical texts extracted from `bible.com`. If data on the desired writing system was not available on `bible.com`, we used data from (Bentz & Ferrer-i-Cancho, 2016), which is also based on biblical texts, instead. Additionally, for Shavian, we extracted the data from `shavian.info/books/`. The texts were preprocessed to remove the punctuation, numbers, and characters that do not belong to the writing system of interest (for instance, Latin characters were removed). The character counts were computed from preprocessed texts and converted to relative frequencies by dividing each count by the sum of counts for the given writing system. Additionally, as the distribution of relative frequencies is highly skewed, these values were log-transformed. This transformation did not affect the results we present here.

The resulting dataset has 27 writing systems. The median corpus size (in characters) is 711,785, with the smallest values for Shavian (97,566 characters) and the largest for Thai (2,942,793 characters). The median number of characters is 42; the writing system with the lowest number of characters is Syriac (22 characters), and the largest writing system is Ethiopic (251 characters). Our dataset consists of four abjads, fourteen abugidas, five alphabets, one featural system, and four syllabaries (1560 characters in total). The geographic distribution of the writing systems in the dataset is shown in Fig. 1:

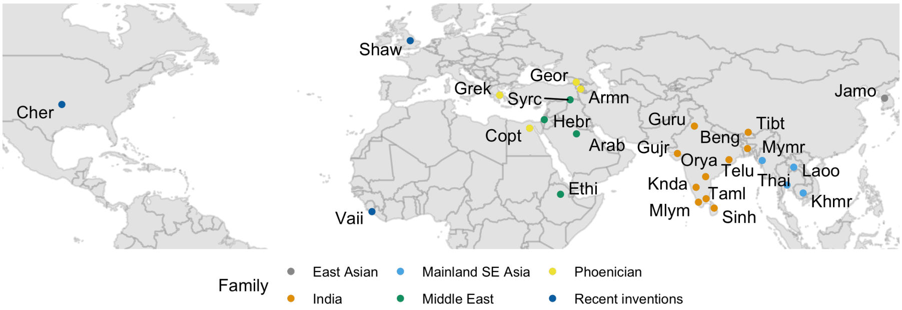


Figure 1. Geographic distribution of the writing systems in the database, annotated with the ISO 15924 codes

## 2.2. *Inclusion criteria*

We included writing systems in our dataset based on several criteria:

1. A writing system was included if it had available Unicode-encoded text files.

2. It is possible to identify one main language for which the writing system was

designed. The Latin and Devanagari writing systems had to be excluded because each of them is used to encode a multiplicity of languages, and each was substantially transformed to encode these languages.

3. The writing system is not combined with other writing systems. For instance, Limbu writing consists of both Devanagari and Limbu characters. Therefore, it was excluded from the sample. However, if the instances of such use are not common, these cases would be kept. For instance, Korean writing today is overwhelmingly based on the Hangul writing system, with only occasional use of Hanja (Mandarin Chinese characters). We focused on analyzing Hangul and disregarded Hanja.
4. Writing systems with less than a hundred thousand characters in the available texts were excluded.

## 3. Analysis

The proposed hypothesis was tested using a mixed-effect linear regression[1] predicting a character's complexity from its relative frequency (fixed effect FREQUENCY) and the writing system to which the character belongs (random effect Writing System). The model has both random slopes and random intercepts for each writing system and was run separately on our algorithmic complexity measure and on our perimetric complexity measure.

First, we measured the null model's Akaike information criterion (AIC). (The null model included only the random effect of Writing Systems.) We compared the null model's AIC with the full model's AIC. The full model included a fixed effect for FREQUENCY and the random effect of Writing Systems, with random slopes for each writing system. If the full model has lower AIC values than the null model (with the conventional threshold being $\Delta\text{AIC} > 2$), that means that the former is more informative than the latter. For perimetric complexity, the $\Delta$AIC value is equal to 172.8, and for algorithmic complexity, this value corresponds to 146.1, meaning that they are both more informative than their respective null models. The $\beta$ coefficients for relative frequency in the perimetric complexity (-2.40, 95% CI: [-1.76, -3.07]) and in the algorithmic complexity models (-26.6 95%CI: [-34.15, -19.17])) are both negative. These values of the coefficients indicate that the higher the frequency is, the less complex is the character, as illustrated in Fig. 2:

---

[1] We used the lme4 R-package to fit our models (Bates, Mächler, Bolker, & Walker, 2014)

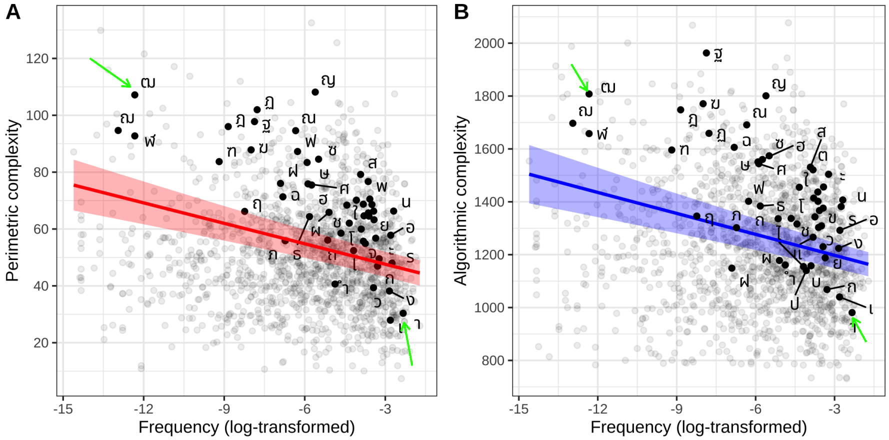


Figure 2. Predictions from perimetric complexity (A) and algorithmic complexity (B) mixed-effect linear regression models for all the scripts combined. Each point corresponds to a unique character. Shaded areas represent the 95% confidence interval for the predictions.

In Fig. 2, each dot represents an individual character. We added Thai characters to each plot for illustrative purposes. The arrows point to the most complex and less complex characters. Additionally, our results suggest that the effects hold for each writing system and are not an artifact from the aggregated data, see Fig. 3:

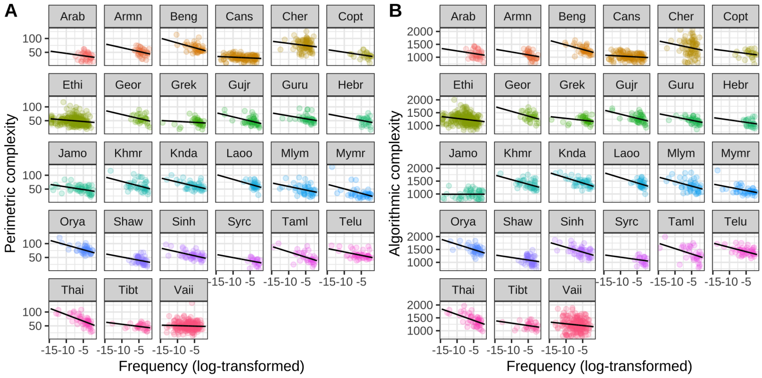


Figure 3. Predictions from perimetric complexity (A) and algorithmic complexity (B) models for individual writing systems. Each point represents an individual character, and each subplot corresponds to an individual writing system (annotated by its ISO 15924 code).

## 4. Conclusions

Using mixed effect linear regression models, we show that Zipf's law of abbreviation holds on all of the individual writing systems in our dataset, not just on the aggregated data taken as a whole, validating our preregistered predictions. These results hold for both our complexity measures and suggest that the law of abbreviation holds in a large variety of writing systems at the level of characters.

Zipf's law of abbreviation was also attested in iterative learning experiments (Kanwal et al., 2017), where the authors suggested that the need for efficient and accurate communication yields the inverse relationship between word frequency and its length. Since our results support the presence of Zipf's law of abbreviation in written communication, this suggests that the results from (Kanwal et al., 2017) can be expanded to graphic communication. Overall, such evidence for Zipf's law of abbreviation in writing supports the idea that optimizing production and reception costs would have been an important factor in the evolution of spoken and graphic communication.

### Supplementary Materials

The preregistration for this study can be found here: `https://osf.io/ydr3n`

A Git-Hub repository will all of the data and code can be found here: `https://github.com/alexeykosh/2021-slojnost-project`

### Acknowledgments

We thank Christian Bentz and Ramon Ferrer-i-Cancho for their helpful advice on this project, and for sharing their data.

### References

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological bulletin*, *53*(6), 452.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bentz, C., & Ferrer-i-Cancho, R. (2016). Zipf's law of abbreviation as a language universal. In *Proceedings of the leiden workshop on capturing phylogenetic algorithms for linguistics* (pp. 1–4).

Chang, L.-Y., Chen, Y.-C., & Perfetti, C. A. (2018). Graphcom: A multidimensional measure of graphic complexity applied to 131 written languages. *Behavior research methods*, *50*(1), 427–449.

Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1560), 267–275.

Ferrer-i-Cancho, R., & Lusseau, D. (2009). Efficient coding in dolphin surface behavioral patterns. *Complexity*, *14*(5), 23–25.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive science*, *31*(6), 961–987.

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52.

Kelly, P., Winters, J., Miton, H., & Morin, O. (2021). The predictable evolution of letter shapes: An emergent script of west africa recapitulates historical change in writing systems. *Current Anthropology*, *62*(6), 000–000.

Miton, H., & Morin, O. (2021). Graphic complexity in writing systems. *Cognition*, *214*, 104771.

Morin, O., Kelly, P., & Winters, J. (2020). Writing, graphic codes, and asynchronous communication. *Topics in cognitive science*, *12*(2), 727–743.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision research*, *46*(28), 4646–4674.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

Rovenchak, A., Mačutek, J., & Riley, C. (2008). *Distribution of complexities in the vai script.*

Rovenchak, A. A., & Vydrin, V. (2010). *Quantitative properties of the nko writing system.*

Semple, S., Hsu, M. J., & Agoramoorthy, G. (2010). Efficiency of coding in macaque vocal communication. *Biology Letters*, *6*(4), 469–471.

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school chinese: Implications for learning to read. *Child development*, *74*(1), 27–47.

Tamariz, M., & Kirby, S. (2015). Culture: copying, compression, and conventionality. *Cognitive science*, *39*(1), 171–183.

Watson, A. B. (2012). *Perimetric complexity of binary digital images: Notes on calculation and relation to visual complexity* (Vol. 14; Tech. Rep.).

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology.* Ravenio Books.

# MODELLING THE EMERGENCE OF VOCAL GROOMING

Tom Kouwenhoven[*1], Ross Towns[1], and Tessa Verhoef[1]

[*]Corresponding Author: t.kouwenhoven@liacs.leidenuniv.nl
[1]Creative Intelligence Lab, Leiden University, Leiden, The Netherlands

The so-called Gossip and Grooming hypothesis (Dunbar, 1998, 2017), posits that spoken language arose as an alternative to allogrooming as a means of maintaining social bonds in growing group sizes among early humans. Though aspects of Dunbar's theoretical framework have been criticised (e.g. Lindenfors, Wartel, and Lind (2021)), the role of language in social bonding is more widely supported (Tomasello, Carpenter, Call, Behne, & Moll, 2005), and is a key component of the musical protolanguage hypothesis (Fitch, 2017). Here we present a multi-agent model based on work by Slingerland, Mulder, Vaart, and Verbrugge (2009), who showed that greater group sizes can stimulate the use of language within a grooming vs gossip paradigm. We update this model by simulating inter-group contact and subjecting agents to selection pressures acting on tolerance for out-group members, drawing on evidence for increased social connectivity in archaeology (Belfer-Cohen & Hovers, 2020) and selection against reactive aggression in early humans (Benítez-Burraco & Progovac, 2020).

## The model

Slingerland et al. (2009) simulated a single group of agents who interact for a number of rounds. Whether this interaction takes the form of one-to-one grooming or one-to-many gossiping is a function of the agents' heritable *gossip probability*. Agents maintain a memory ($M$) of social interactions that are acquired by participating in, observing, or gossiping about an event. Selection is based on an equal combination of social fitness ($f_{social} = 5 \times \sum_{x=0}^{E_{groom}} \frac{1}{p_x - 1} + 4 \times \sum_{x=0}^{E_{gossip}} \frac{1}{p_x - 1}$ , where $p_x$ is the number of agents in a social event $x$ out of all events $E$) and information fitness ($f_{info} = M^2$) determines which agents reproduce offspring who inherit mutated preferences from their parents.

Our first addition involves investigating group dynamics by dividing the population into multiple groups. The likelihood that an agent interacts with an in- or out-group individual is a function of its heritable *tolerance* preference. Secondly, a recalculation of data (Nakamura, 2000, 2003) and new observations of bonobos and chimpanzees in the wild (Girard-Buttoz et al., 2020) reveal that, contrary to what was originally thought, social grooming is not strictly dyadic and one-way but is often polyadic and mutual. In line with these findings, gossip and grooming are both one-to-many in the updated model. We investigate if, under these circumstances, gossip still becomes dominant and what the interaction between group dynamics and tolerance is.

Figure 1 shows that a preference for gossip still evolves, in addition to clearly distinct patterns for differently evolved group sizes: agents of large groups develop a preference for gossip with in-group members (low tolerance), while agents of smaller groups rely on both high tolerance and gossip probability.

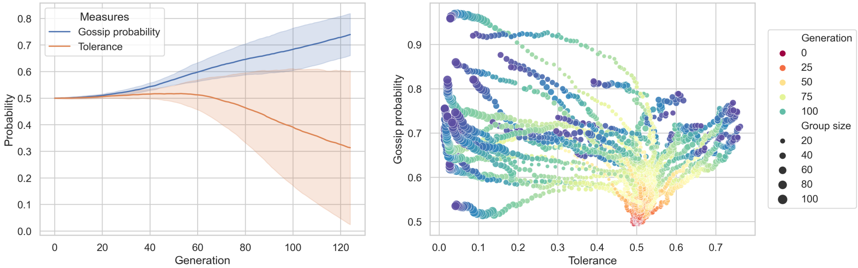


Figure 1. The evolution of tolerance and gossip probability over generations (left). The dynamic between average group size, generation, tolerance, and gossip probability on population level (right). Shaded regions are SD over 40 simulations of 125 generations initialised with 100 agents in 20 groups.

## Discussion

We show that vocal grooming (gossip) can emerge as the dominant bonding strategy even when the assumption that clique sizes are different for gossip and grooming are relaxed. While Dunbar (2017) emphasises the role of time constraints in restricting the number of social bonds that can be maintained through one-to-one grooming, as a constraint of group size, our model suggests that the benefit of having information about social events mainly drives the emergence of gossip. Moreover, the results show that the evolution of tolerance is influenced by group structure in the community. Similarly, such ecological factors have been suggested to play a role in inter-group tolerance for primates (Lucchesi et al., 2020). For humans, however, the evolution of inter-group tolerance was likely affected by many other factors including mating patterns and food availability (Spikins, French, John-Wood, & Dytham, 2021). Notably, the Human Self-Domestication hypothesis (Hare, 2017) posits that direct selection against reactive aggression increased tolerance in humans, which may have facilitated more inter-group encounters (Benítez-Burraco & Progovac, 2020). The precise interactions between different factors affecting inter-group tolerance, population dynamics and how these shape the evolution of language is what we propose to investigate further.

## References

Belfer-Cohen, A., & Hovers, E. (2020). Prehistoric perspectives on “others” and “strangers”. *Frontiers in Psychology*, *10*, 3063.

Benítez-Burraco, A., & Progovac, L. (2020). A four-stage model for language evolution under the effects of human self-domestication. *Language & Communication*, *73*, 1-17.

Dunbar, R. I. (2017). Group size, vocal grooming and the origins of language. *Psychonomic bulletin & review*, *24*(1), 209–212.

Dunbar, R. I. M. (1998). *Grooming, gossip, and the evolution of language.* Harvard University Press.

Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic bulletin & review*, *24*(1), 3–33.

Girard-Buttoz, C., Surbeck, M., Samuni, L., Boesch, C., Fruth, B., Crockford, C., Hohmann, G., & Wittig, R. M. (2020). Variable use of polyadic grooming and its effect on access to social partners in wild chimpanzees and bonobos. *Animal Behaviour*, *168*, 211–224.

Hare, B. (2017). Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality . *Annual Review of Psychology*, *68*(1), 155–186.

Lindenfors, P., Wartel, A., & Lind, J. (2021). ‘dunbar’s number’deconstructed. *Biology Letters*, *17*(5), 20210158.

Lucchesi, S., Cheng, L., Janmaat, K., Mundry, R., Pisor, A., & Surbeck, M. (2020). Beyond the group: how food, mates, and group size influence intergroup encounters in wild bonobos. *Behavioral Ecology*, *31*(2), 519–532.

Nakamura, M. (2000). Is human conversation more efficient than chimpanzee grooming? *Human Nature*, *11*(3), 281–297.

Nakamura, M. (2003). ‘gatherings’ of social grooming among wild chimpanzees: implications for evolution of sociality. *Journal of human evolution*, *44*(1), 59–71.

Slingerland, I., Mulder, M., Vaart, E. van der, & Verbrugge, R. (2009). A multi-agent systems approach to gossip and the evolution of language. In *Proceedings of the 31st annual meeting of the cognitive science society (cogsci’09)* (pp. 1609–1614).

Spikins, P., French, J. C., John-Wood, S., & Dytham, C. (2021). Theoretical and methodological approaches to ecological changes, social behaviour and human intergroup tolerance 300,000 to 30,000 bp. *Journal of archaeological method and theory*, *28*(1), 53–75.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(5), 675–691.

# SEMIOTIC ARCHAEOLOGY: CORRELATED EVOLUTIONARY STAGES OF LANGUAGE, ART AND TOOLS

ALEXANDRA R. KRATSCHMER*[1] and JACOB WAMBERG[2]

*Corresponding Author: romak@cc.au.dk

[1]Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, Denmark
[2]Department of Art History, Aesthetics and Culture and Museology, Aarhus University, Denmark

Bringing together the archaeological record from art history and tool production, this paper delineates four different stages of semiotic and hence communicative abilities for hominin evolution from the primate stage to modern *homo sapiens*: (1) reactive-mimetic; (2) proactive-mimetic; (3) hyperindexical; and (4) externalized iconic and symbolic. These evolutionary stages presuppose a gradual emancipation of a virtual sphere detached from nature and a corresponding giving over of hominin evolution to culturally rather than genetically determined conditions. This transition comprises a spiraling turn in the Lower Paleolithic, in which the phylogenetic lifeworld emerges as an enlargement of the intimate space of postnatal ontogenesis, what we term a macro-cocoon. Although we hold that any hominin line would possibly follow this route of development, either independently or in interaction with other contemporary lines, only *Homo Sapiens* reached the fourth stage.

## 1. Introduction: an interdisciplinary approach

This paper presents a transdisciplinary model of human evolution, linking linguistics with art history and prehistoric archeology. Its approach is systems-oriented, suggesting that humanoid mental capacities have evolved in increasing interaction with diverse forms of cultural practices. Because of the close interaction between "outer" cultural practices and their imprints in "inner" brain disposition, we must suppose certain stages of human evolution, in which archaeological traces of semiotic activities and utilitarian artifacts could be seen as evidence of stages of certain brain dispositions, spanning cognition, emotions and religious beliefs. We suggest a preliminary model with four basic steps, whose ground tendencies are supposed to be applicable to any hominin line, living or extinct. Although any hominin line would possibly follow this route of development, either independently or in interaction with other contemporary lines, only *Homo sapiens* reached the fourth stage. In our concluding synoptic table, we exemplify the different stages by samples from the currently available archaeological record, but without claiming them to be exhaustive.

Our guiding presupposition is that the advancement of hominin culture creates a virtual niche, or sphere, that is increasingly autonomous from the living and surviving conditions offered by the natural habitat (Harari [2011] 2014), and that this involves a corresponding autonomization of semiotic signs in relation to their referents. It is indeed the crystallization of such a relatively autonomized virtual sphere that pauses the genetically based anatomical evolution of the human species. Human evolution is now predominantly transferred – sublimated – to the domain of social learning, i.e. accumulative culture. Here, anatomical changes only happen after birth in the form of micro-changes in cortical morphology and diverse bodily techniques (Deacon, 2010; Krubitzer & Prescott, 2019).

However, the mature stage of this sphere, presenting the first externalized icons and symbols, was only reached after a long evolutionary process that could be compared to a spiraling turn. In total, we distinguish the following four major evolutionary stages: (1) *reactive-mimetic*: a stage, shared with other primates, as well as dolphins and certain corvids, in which the early hominin can recognize its mirror image as referring to itself; (2) *proactive-mimetic* (beginning 2,5 mio BP): the earliest specifically humanoid stage marked by stone tools, the ability to project iconic shapes onto pre-existing natural forms, and mimetic body language; (3) *hyperindexical* (beginning ca. 700-500,000 BP): a stage in which the phylogenetic life world at large becomes modeled on the intimate space of early postnatal ontogenesis, resulting in an animated ecology with the first thoroughly externalized semiosis: a stylized reactualization of indexical communication, including proto-symbolic speech (musical proto-language) and nonfigurative pictorial patterns; and (4) *externalized iconic and symbolic* (beginning ca. 100,000-40,000 BP), marked by representational pictures and symbolic language.

## 2. The four evolutionary stages: semiosis of art, toolmaking and language

The initial stage (1) of hominin semiosis, *reactive-mimetic*, has to be the one present in contemporary great apes (Anderson & Gallup, 2011), which we assume was prevalent in both ancestral non-hominin primates and early hominins – the ability to recognize oneself in a mirror (e.g. of still water). This ability relies on the immediate indexical association between the self and its reflection. Although the hominin mind hereby identifies itself with something outside the body, this identification does not yet presuppose a consciousness of the distinction between body and mirror reflection but merely their close association. As is also the case with contemporary apes, this ability to self-reflect is allied with an ability to use stone tools (Kortland, 2013) and to communicate by vocalizations and gestures, with a predominance of the latter in close-range communication, characterized by intentionality (deictic) reference, iconicity (pantomime), and turn-taking, i.e. communication structured between dialogue partners (Fröhlich et al. 2019).

The second stage, the *proactive-mimetic* one (2), covers the two tool industry periods of the Oldowan and the Acheulean. Its initiation is exemplified by the South African Makapansgat Pebble, a hominid face-like, but nature-created, stone that 3-2,5 mio BP apparently was retrieved and transported from its finding site to their habitation 50 miles away by supposedly *Australopithecines*. This effort could be seen as a manifestation of the general idea of 'human face', in Peirce's categories of a so-called legisign, "a law that is a sign." However, at this stage, the representation is established only through reaction to something already given, rather than production (Watson, 2015 as quoted in Hampton, s.a.). What is specifically hominin and not found in apes is the ability to understand the double nature of the image as a sign that is both similar and dissimilar to the thing it represents. Probably the oldest documented hominid stone tool production, consisting of stone tools coarsely chipped at the edges to yield a cutting edge (de la Torre, 2011), the East African Oldowan culture by *Homo habilis*, from ca. 2,6-1,7 mio BP, represents the first level of tool production linked to this mental level of image recognition (Sernaw et al. 2003). The next level of stone tools within this stage is the Acheulian type (hand axes and cleavers): bifacially knapped, symmetrical artifacts, produced by the homo lineages *Erectus* and *Heidelbergensis* (Shipton & Nielsen, 2018). The earliest record so far dates back to 1.76 mio BP (Lepre et al., 2011), and the industry remained productive until about 130,000 BP. Peterson et al. (2018) specify the existence of the Acheulean hand axe as a type (category/model) to be realized in form of a token (the concrete handaxe) in the terms of the Peircian sign types of respectively "legisign" (type/an idea) and "sinsign" (token/an actual singular thing) (Bergmann et al., 2003f.). We hold that, while interior legisign and exterior sinsign are in fact highly detached at this evolutionary stage, the sign user does not experience them as separate. At stage (2), intra- and intergenerational technology was therefore still transferable through mimetic-imitative embodied learning alone.

To approach the complex *hyperindexical* stage (3), we should understand it as a phase that moves the expanding world of inner representations closer to the outer world of objects through dense indexical connections. It could be specified as what Tomlinson (2017: 13) calls "hyperindexicality," that is, the "systematic and hierarchical arrangement of indexes in relation to one another, which brings them close to one of the characteristic features of the symbol." How this bridge-building between inner and outer worlds is realized could be illuminated by considering the further evolution of hominins as an exposure of a now more vulnerable organism towards its accordingly more protective environment. For a crucial price for the evolutionary rise of complex brains and sensoria seems to be an intensification of the initial protection of individual organisms, which is underway throughout the evolution of eukariotic organisms: the ontogenetic protective worlds of eggs, wombs, nests, and warm spaces for suckling. Hominins are located at a particularly intricate crossing point between rising autonomization of the brain and this accompanying protection of embryos and infants, since now this protective cocoon not only restricts itself to the ontogenetic evolution of womb and suckling environment. Rather, in a transferal from onto- to phylogenesis, it seems to be projected onto the physical environment at large in a process, which the German philosopher of culture, Peter Sloterdijk (1999), terms *Blasen* (blowing). Corresponding with this giving-over of protection to the macro-cocoon of the surrounding world, the hominin body seems to become increasingly fragile and open in its naked state. Prematurely born and continuing its neoteny, the furless vulnerable state of young mammals into adulthood, the naked human ape (Morris, 1976) only becomes completed through wrapping itself in new artificial furring: the macro-cocoon of semiotics and technology (Gehlen [1940] 2016). It is this process of self-domestication (Deacon, 2010) that generates the hyperindexical stage, in which inner signs are connected to the outer world of objects and events through intense indexical communication.

Although evolutionary linguistics seems to lack a systematized model of this process of self-domestication, Darwin's original idea about a "musical proto-language" has lately gained renewed scholarly interest and might be a well-suited explanatory candidate for accounting for the communicative underpinnings of this process: an indexical form of proto-symbolic communication, in which sounds are stylized according to intensity, rhythm and pitch (Ravignani & de Boer, 2021). As suggested by Michael Tomasello (2008), this sort of proto-symbolic language is very similar to the nonlinguistic communication between early infants and caretakers. To further elaborate this connection between onto- and phylogenetic development we should consult the paleo-art historian Ellen Dissanayake (2000) who traces the origins of visual art to exactly a phylogenetic upscaling of the intimate indexical communication between parent and infant.

In terms of visual art, the pre-symbolic indexical patterns of parent-infant communication are confirmed by the surviving remnants of Lower Paleolithic art, from *Homo heidelbergensis* to *Homo neanderthalensis*. Their exclusively non-figurative patterns, such as parallel lines or criss-crossings, could thus be conceived of as ritualistic traces, imprints of performative actions, whose significance is retained in indexical form. The main surviving evidence includes the *Pseudodon Shell DUB 1006-fL* (540,000-430,000 BP) from Trinil, Java (Joordens et al., 2015), the elephant bone from Bilzingsleben, Germany (400,000-350,000 BP) (Mania & Mania, 1988), the ochre and ostrich eggshell fragments from the South African Blombos Cave and Diepkloof Rock Shelter (100,000 BP) (Tylén at al., 2020), and the criss-cross patterns engraved by Neanderthals in the rocky walls of the Gorham Cave, Gibraltar (ca. 39,000 BCW) (Callaway, 2014). The same semiotic mechanisms must have applied in the case of the Schöningen wooden hunting spears (nine spears, one lance, a double pointed stick, and a burnt stick) from 300,000 BP, made by *Homo heidelbergensis* or early *Neanderthalensis* (Schoch et al., 2015). The cultural niche, where abstract models (types) and concrete tokens must coexist in the minds of the artisans, allows for the passage from indexical to symbolic thinking. On the practical level of communication, the indexical event of namegiving of concrete tokens (living or non-living entities) precedes the symbolic event of namegiving of categories/types (legisigns), and is supposedly found at this cultural state (Deacon 1997).

Stage (4) is finally characterized by externalized icons (representational pictures) and externalized symbols (spoken words), with the latter possibly preceding the former, at least if we keep to what the currently available archaeological record allows us to infer. The externalization of mammalian and avian fauna in the cave paintings of Spain and France, and on the Indonesian island of Sulawesi, created during a period from ca. 45,500 to 12,000 BP (García-Diez et al., 2013; Ducasse & Langlais, 2019), were probably dependent on a beginning division of the Lower Paleolithic macro-cocoon. Because humans now felt separated from an emergent maternal underworld, a bridge was constructed in the form of pictorial animals that seemed autochthonously born through the cave walls, the membrane between the other world and ours (Clottes & Lewis-Williams, [1996] 1998). Such externalized pictorial signs thereby made remote realities of all kinds accessible, including those of imagined abstractions. If we now look at the archaeological evidence for bow-and-arrow technology, the complexity of its complementary design (hafting of different components with several-component glues, etc.) demands communication of highly abstract information, which can only be performed via a similarly externalized symbolic language system (Wadley, 2010; Lombard & Haidle, 2012). The emergence of such a system is hence a prerequisite for this cultural stage, whose archeological record can be traced back to 70,000 BCW.

In the following synoptic *Table 1,* we will, in conclusion, recapitulate the steps of this chronology, while emphasizing that the datings must be taken as approximative and veiling the plausible assumption that there has been considerable overlap in time as well as depending on space.

Table 1. Conservative evolutionary chronology of semiotic stages, art production, tool making and communication in hominins (approximative and simplified).

| *Dating* | *Semiotic stages* | *Art* | *Tools* | *Communication* |
|---|---|---|---|---|
| primate | reactive-mimetic | recognition of own mirror image (pre-stage) | stones and sticks | gestural, vocal |
| 3-2,5 mio | proactive-mimetic; legisigns and sinsigns | Makapansgat pebble | | vocal, gestural (mimetic) |
| 2,6 mio | | | Oldowan | |
| 1,76 mio – 130.000 | | | Acheulean | |
| 540-430.000 | hyperindexical | Trinil Shell | | namegiving of tokens |
| 400-350.000 | | Bilzingsleben Bone | | |
| 300.000 | | | Schöningen spears | |
| 100.000 | | Blombos, Diepkloof engravings | | |
| 70.000 | externalized symbolic and iconic | | bow & arrow | symbolic language<br>(namegiving of categories/types) |
| 45.000 | | Sulawesi paintings | | |
| 20.000 | | Lascaux paintings | | |

## References

Anderson, J.R.. & Gallup, G.G. Jr. (2011). Which Primates Recognize Themselves in Mirrors? *PLoS Biol, 9(3)*, e1001024.

Barham, L. & Everett, D. (2021). Semiotics and the origin of language in the Lower Paleolithic. *Journal of Archaeological Method and Theory, 28 (2),* 535-579.

Bergman, M., Paavola, S. & Queiroz, J. (2003f.). *Commens. Digital Companion to C. S. Peirce.* http://www.commens.org/home

Callaway, E. (2014). Neanderthals made some of Europe's oldest art. *Nature.* https://doi.org/10.1038/nature.2014.15805

Clottes, J. & Lewis-Williams, D. ([1996] 1998). *The Shamans of Prehistory: Trance and Magic in the Painted Caves*, trans. Sophie Hawkes (New York: Harry N. Abrams).

Deacon, T. W. (1997). *The symbolic species. The co-evolution of Language and the brain*. New York and London: The W. W. Norton Company.

Deacon, T. W. (2010). *Language and complexity: Evolution inside out.* Lecture held at the 37th International Systemic Functional Congress, https://www.youtube.com/watch?v=OT-zZ0PMqgI

Dissanayake, E. (2002). *Art and Intimacy: How the Arts Began*. Seattle and London: University of Washington Press.

Ducasse, S. & Langlais, M. (2019). Twenty years on, a new date with Lascaux. Reassessing the chronology of the cave's Paleolithic occupations through new 14C AMS dating. *Revue d'Archéologie Préhistorique, 30 (1),* 130-147.

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T. & van Schaik, C. P. (2019). Multimodal communication and language origins: integrating gestures and vocalizations. In *Biological Reviews, 94*, 1809-1829.

García-Diez, M., Hoffmann, D. L., Zilhão, J., de las Heras, C., Lasheras, J.A., Montes, R. & Pike, A.W.G. (2013). Uranium series dating reveals a long sequence of rock art at Altamira Cave (Santillana del Mar, Cantabria). *Journal of Archaeological Science, 40 (11),* 4098-4106.

Gehlen, Arnold. *Der Mensch: Seine Natur und seine Stellung in der Welt* (Frankfurt a.M.: Vittorio Klostermann, [1940] 2016).

Hampton, J. G. (s.a.). Contemporary Rock Art. In Hampton, J. G. (s.s.). *Rocks, Stones Dust, Contemporary Art Devoted to Rocks and Their Relations*. Online Exhibition Catalog. The Justina M. Barnicke Gallery and the University of Toronto Art Centre, *28*, 1-22.

Harari, Y. N. ([2011] 2014). *Sapiens: A Brief History of Humankind*, trans. P.P.P.S.F. Ahmed. New York: Harper.

Joordens, J. C. A., d'Errico, F., Wesselingh, F. P., Munro, S., de Vos, J., Wallinga, J., Ankjærgaard, C., Reimann, T., Wijbrans, J. R., Kuiper, K. F. & Mücher, H. J. (2015). Homo erectus at Trinil on Java used shells for tool production and engraving. *Nature, 518 (7538)*, 228-21.

Lombard, M., & Haidle, M. N. (2012). Thinking a Bow-and-Arrow Set: Cognitive Implications of Middle Stone Age Bow and Stone-tipped Arrow Technology. *Cambridge Archaeological Journal, 22 (2)*, 237-264.

Mania, D. & Mania, U. (1988). Deliberate engravings on bone artifacts of Homo Erectus. *Rock Art Research, 5*, 91-97.

Morris, D. (1967). *The Naked Ape.* London: Corgi Books.

Peterson, J.V., Thornburg, A.M., Kissel, M., Ball, K. & Fuentes, A. (2018). Semiotic Mechanisms Underlying Niche Construction. *Biosemiotics, 11*, 181-198.

Pike, A. W. G., Hoffmann, D. L., García-Diez, M., Pettitt, P. B., Alcolea, J., De Balbín, R., González-Sainz, C., de las Heras, C., Lasheras, J. A., Montes, R. & Zilhão, J. (2012). U-Series Dating of Paleolithic Art in 11 Caves in Spain. *Science, 336 (6087),* 1409-1413.

Ravignani, A., & de Boer, B. (2021). Joint origins of speech and music: testing evolutionary hypotheses on modern humans. *Semiotica, 239*, 169-176.

Schoch, W. H., Bigga, G., Böhner, U., Richter, P. & Terberger, T. (2015). New insights on the wooden weapons from the Paleolithic site of Schöningen. *Journal of Human Evolution, 89*, 214-225.

Semaw, S., Rogers, M. J., Quade, J., Renne, P. R., Butler, R. F., Domínguez-Rodrigo, M., Stout, D., Hart, W. S., Pickering, T. & Simpson, S. W. (2003). 2.6-Million- year-old stone tools and associated bones from OGS-6 and OGS-7, Gona, Afar, Ethiopia. *J. of Human Evolution, 45 (2),* 169–177.

Shipton C., Nielsen M. (2018) The Acquisition of Biface Knapping Skill in the Acheulean. In: L. D. Di Paolo, F. Di Vincenzo & F. De Petrillo (Eds.). *Evolution of Primate Social Cognition. Interdisciplinary Evolution Research* (pp. 283-297). Vol 5. New York: Springer.

Sloterdijk, P. (1998). *Sphären I: Blasen*. Frankfurt a.M.: Suhrkamp.

Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. *Child Development, 57(6),* 1454-1463. https://doi.org/10.2307/1130423

Tomasello, M. (2008) *Origins of Human Communication*. Cambridge, MA: MIT Press.

Tomlinson, G. (2017). Semiotic Epicycles and Emergent Thresholds in Human Evolution. *Glass-Bead 1*, 1-14.

Wadley, Lynn. (2010). Compound-adhesive manufacture as a behavioral proxy for complex cognition in the Middle Stone Age. *Current Anthropology, 51 (1),* 111–19.

# NEURAL CORRELATES OF SPEECH AND NON-SPEECH PRODUCTION IN CHILDHOOD

LANCHEROS, M.*[1], ATANASOVA, T. [1], LAGANARO, M.[1]

*Corresponding Author: monica.lancherospompeyo@unige.ch
[1] Faculty of Psychology and Educational Science, University of Geneva, Geneva, Switzerland

Speaking is a complex behavior at the intersection between cognitive and motor processes that implies the coordination of multiple systems and of multiple vocal-tract muscles. The uniquely-human ability to speak indicates that some mechanisms underlying speech production have diverged from those of our prelinguistic ancestors. However, the basic capacity to vocally produce sounds is shared with a surprisingly diverse group of species. In fact, the oromotor structures in charge of articulating speech are not only responsible of the production of speech targets but they also generate other sounded orofacial movements, such as lip smacking or tongue clicks. Although the brain dynamics underlying the production of speech and non-speech gestures in adults are shown to differ (Lancheros, Jouen and Laganaro, 2020) the issue is then when in ontogeny the production of speech differs from that of non-speech. In the present study we explore this question, with the hypothesis of a specialized motor control system for speech arising only after some years of full practice of speech. Since the production of speech starts early after birth and its motor entrainment extends over more than a decade (Ackermann, Hage & Ziegler, 2014), it is likely that a gradual specialization for speech emerges during childhood as a consequence of its progressively extensive production. However, given that the phonetic inventory is not achieved before the age of seven, a distinct specialized speech neural architecture might be established only some years ahead (i.e. >10 y.o.).

In the present study participants included fifteen French-native 7-to-9-year-old children. They were asked to produce 20 sounded non-speech gestures, 20 high frequency syllables and 20 low frequency syllables, paired on type and mode of articulation of the gesture onset. Stimuli were presented throughout video clips in a task targeting motor encoding processes: a delayed production task combined with an articulatory suppression task. Their brain activity was in parallel recorded with electroencephalography (EEG) and event-related potential (ERP) analyses based on microstates were performed on the time window preceding the vocal onset.

Behavioral results showed that children were less precise when producing low frequency syllables as compared to non-speech. Concerning production latencies, non-speech productions were initialized slower than both types of speech stimuli. The ERP data aligned to the vocal onset showed activation of the same topographical maps preceding the production of speech and of non-speech sequences, indicating the recruitment of the same neural networks for non-speech and speech. Additionally, the temporal distribution of the same recruited brain circuits did not differ between the two oromotor behaviors, suggesting similar activation dynamics for speech and non-speech.

Considering that adults do show different brain activation patterns for speech and non-speech (Lancheros et al. 2020) with the exact same experimental setting, results of the present study seem to support the hypothesis that a specialized neural circuit for speech is not established yet in children aged 7-9 years. Since a distinct neural representation for speech has been proposed to arise from its extensive motor learning, it might be the case that children at those ages have not practiced enough this oromotor behavior and thus, speech is not yet a well-established overlearned motor activity with its specialized brain circuits. Unquestionably, speech does not only depend on the amount of rehearsal it is subjected but it also relies on cognitive, linguistic and motor developments occurring during childhood, which might also explain why a specialized neural circuity for speech is not yet mature, or at least it is not neurally evident, in 7-to-9-year-olds.

## References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences, 37*(6), 529-546.

Lancheros, M., Jouen, A., & Laganaro, M. (2020). Neural dynamics of speech and non-speech motor planning. *Brain and language, 203*, 104742.

# THE IMPORTANCE OF THE FUNDAMENTAL FREQUENCY IN CHIMPANZEES VOCAL COMMUNICATION

Marion Laporte*[1,2], Louis-Jean Boë [2,3] Nicolas Audebert [4]

*Corresponding Author: marion.laporte@mnhn.com
[1] UMR 7194, PaleoFED, Muséum National d'Histoire Naturelle, Paris, France
[2] Institut des Sciences du Calcul et des Données, Sorbonne Université, Paris, France
[3] GIPSA Lab, CNRS, Grenoble, France
[4] CEDRIC (EA4629), Conservatoire National des Arts et Métiers, HESAM Université, Paris, France

In the quest for understanding the evolution of human language, primates have often been taken as an approximation of the human ancestral condition despite their equally long evolutionary path. Recent studies have shown that several primates, including chimpanzees, have the capacity to produce vowel-like vocalizations, despite a high larynx (Boë et al., 2017; Fitch et al., 2016; Grawunder et al., 2022). However, many of these primates live in complex social systems in environments where the visibility is low and where they need to communicate over long distances. In these cases, the vocal signals produced involve high intensity signals associated to high fundamental frequencies (*fo*) that are better suited for long-distance propagation as they maximize the sound power radiation (Titze & Palaparthi, 2018). This is especially the case for chimpanzees, who live in complex fission-fusion societies in forests or savannah habitats (Goodall, 1986).

Here, we argue that these long-distance vocalizations which constitute two-thirds of all of male chimpanzees' vocal communication (Arcadi, 2000) cannot always provide complete formantic information or contrast, as is the case for human vowels which are produced with low fundamental frequencies. One way around this problem is to look at the fundamental frequency modulation, which is equally well conserved over long distances and generally robust against surrounding noise. While the chimpanzee vocal repertoire has been extensively studied over the years, no study has yet looked at the *fo* modulation per se and its importance in the vocal repertoire of chimpanzees. We aim to evaluate this by determining a vocal space of the entire vocal repertoire through different yet complementary machine learning approaches using the *fo* modulation and Mel-frequency cepstral coefficients (MFCCs) that are commonly used in human speech recognition and take into account all the spectral information. We manually extracted the temporal evolution of the *fo* for N=+6000 vocal units recorded from wild Ugandan chimpanzees (11 adult males and 12 adult females). This allowed us to treat a vocalization as a univariate time series, making them suitable for further analysis. We tested both *supervised* classification methods (Support Vector Machines (SVM) (Schölkopf & Smola, 2001) using the GAK kernel (Cuturi, 2011)) and trained an auto-encoder neural network on the frequential time series to perform a non-linear dimension reduction for *unsupervised* classification. The interesting aspect of working without supervision is that it does not rely on an a priori manual classification of vocalizations into different groups and helps to identify patterns in the data without human bias (t-SNE). We compared the results obtained on the *fo* profiles with equivalent classification and dimension reduction techniques applied to the spectral information *(MFCCs)*, similarly to how previous works performed data analysis on primate vocalisations (Wadewitz et al., 2015) and CNN 2D typically used for image classification but with good performances on spectrogram classification (Hershey et al., 2017).

Our results show that the modulation of the fundamental frequency is an important factor in differentiating the vocalisations of the chimpanzee repertoire, despite an important gradation between calls. An analysis of the *fo* modulation alone obtains good results. The classification does not improve with the MFCC approach, but gains a finer-grained categorization with the CNN 2D approach. We present clustering provided through t-SNE visualization which permits to better understand the links between the different vocalisations and the importance of their gradation.

We examine these results from an evolutionary perspective and discuss the importance of the *fo* modulation in chimpanzee vocal communication compared to human speech production.

## Acknowledgements

This work is part of the Apesvoice project, funded by Sorbonne University (Emergence 2021-2022) and the ISCD as part of the *Origins of Speech* team. We would like to thank Amélie Vialet and Pascal Perrier for their comments.

## References

Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., Becker, Y., Rey, A., & Fagot, J. (2017). Evidence of a Vocalic Proto-System in the Baboon (Papio papio) Suggests Pre-Hominin Speech Precursors. PLOS ONE, 12(1), e0169321. https://doi.org/10.1371/journal.pone.0169321

Cuturi, M. (2011). Fast Global Alignment Kernels. In Proceedings of the 28th international conference on machine learning (ICML-11), pp. 929–936.

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. Science Advances, 2(12), e1600723. https://doi.org/10.1126/sciadv.1600723

Goodall, J. (1986). The chimpanzees of Gombe: Patterns of behavior. Cambridge: Harvard University Press.

Grawunder, S., Uomini, N., Samuni, L., Bortolato, T., Girard-Buttoz, C., Wittig, R. M., & Crockford, C. (2022). Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. Philosophical Transactions of the Royal Society B: Biological Sciences, 377(1841), 20200455. https://doi.org/10.1098/rstb.2020.0455

Hershey, S. et al. (2017) « CNN architectures for large-scale audio classification ». In : 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). p. 131-135.

Schölkopf, B., & Smola, A. J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

Titze, I., & Palaparthi, A. (2018). Radiation efficiency for long-range vocal communication in mammals and birds. The Journal of the Acoustical Society of America, 143(5), 2813–2824. https://doi.org/10.1121/1.5034768

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing Vocal Repertoires—Hard vs. Soft Classification Approaches. PLOS ONE, 10(4), e0125785. https://doi.org/10.1371/journal.pone.0125785

# WORD SEGMENTATION IS FACILITATED IN ZIPFIAN DISTRIBUTIONS

Lavi-Rotbain, O. *1, Arnon, I. 1

*Corresponding Author: orilavirotbain@gmail.com
1Psychology Department, Hebrew University, Jerusalem, ISRAEL

While languages differ from one another in many respects, they share certain commonalties: these can provide insight on our shared cognition and the way it may impact language structure. Here, we focus on one of the striking commonalities between languages, the way word frequencies are distributed. Across languages, words follow a Zipfian (or near-Zipfian) distribution, showing a power law relation between a word's frequency and its rank (Piantadosi, 2014; Zipf, 1949). The source of such distributions in language is debated (e.g., Chater & Brown, 1999), but they have been proposed to reflect foundational aspects of human cognition and/or communication (e.g., Ferrer i Cancho & Sole, 2003; Ferrer i Cancho et al. 2020; Manin, 2008). Regardless of their source, their recurrence in language may have advantages for learning: skewed distributions may facilitate various aspects of language. Such individual biases could be amplified over time, creating pressure to maintain skewed distributions (as has been argued for other linguistic domains, e.g., Kirby et al. 2008).

In this paper, we explore the learnability consequences of Zipfian distributions for word segmentation, a crucial aspect of early language acquisition. Word segmentation has been studied extensively in the lab using artificial language learning tasks (Saffran et al. 1996), but learners are usually presented with a uniform learning environment where each novel word appears equally often. The few studies that examined learning from Zipfian distributions suggest they are beneficial for word segmentation (Kurumada et al. 2013; Meylan et al. 2012), but the extent and generality of this effect is unclear. More importantly, we do not know *what* about Zipfian distributions impacts learning.

Here, we propose and test the prediction that Zipfian distributions are facilitative because of their lower unigram entropy. We start by quantifying unigram entropy (using efficiency, a normalized entropy measure, see Eq. 1) in child-directed speech – children's actual learning environment - across 15 languages (following Bentz et al. 2017 who did so for adult-to-adult speech). We find that efficiency spans a surprisingly narrow range across languages (range=0.6-0.7, mean=0.63, SD=0.03). We then test the impact of those values on learning by manipulating unigram entropy and distribution shape using a classic artificial word segmentation paradigm. We compare performance at three efficiency levels: maximal (uniform distribution), reduced (skewed distribution, with lower unigram entropy than the uniform but higher than natural language), and language-like (within the range found for natural language). We find that word segmentation in adults is uniquely facilitated in language-like efficiency overall (see Fig. 1a), and for the low frequency words (see Fig. 1b), but does not improve in reduced efficiency. We find similar results in a second study using a differently skewed distribution with similar efficiency.

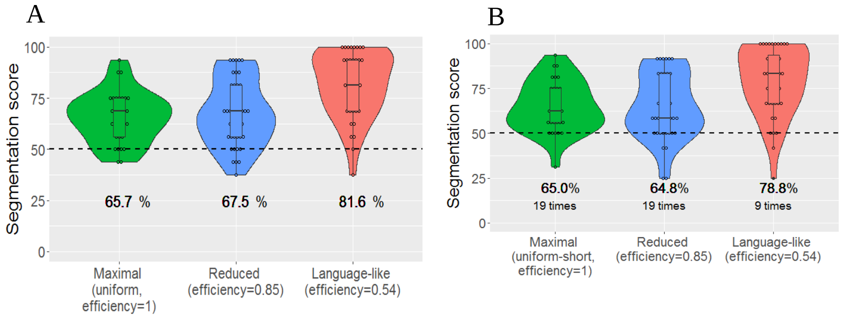


Figure 1. Adults' segmentation scores across conditions (a), and for low frequency words (b).

These findings show that child-directed speech has similar unigram entropy across languages, and that these values are uniquely facilitative for word segmentation. We discuss the possible role of learnability pressures in the emergence of such distributions.

## References

Ferrer i Cancho, R., & Sole, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*(3), 788–791.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681-10686.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, *21*(5), 1112–1130. Retrieved from

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928.

Zipf, G. K. (1949). Human behavior and the principle of least effort. In *Human* *behavior and the principle of least effort.* Oxford, England

# HOW THE STRUCTURES OF UMWELTS AND SIGNS ARE RELATED

Eric Leonardis*[*1], Arturs Semenuks*[1], Greyson Sims[2], and Seana Coulson[1]

[*]Corresponding Authors: ejleonar@ucsd.edu; asemenuk@ucsd.edu
[1]Department of Cognitive Science, UCSD, La Jolla, USA
[2]Department of Philosophy, UCSD, La Jolla, USA

Why do humans use symbols more than any other animal (Deacon, 1998)? Here we argue that part of the explanation might lie in the effect certain properties of the human sensory world have on language.

Following Peirce (1955), semioticians usually distinguish three kinds of signs – indices, icons, and symbols – that are often treated as discrete, separate classes. However, analyzing the structure of sign types using conceptual blending theory (CBT: Fauconnier and Turner (2008)) allows us to establish a continuity between them. In CBT, mental spaces are integrated by a projection of those spaces as inputs onto an emergent blend space, which develops structure beyond the input spaces alone. Conceptual blends could be ordered on a cline based on how incongruous the structures of their input spaces are. Human thought frequently features blends with input mental spaces whose structures are in 'fundamental conflict' (Turner, 2014). Signs can be analyzed as blends of the signifier (symbol or other representation) and the signified (referent) (Peirce, 1955). Indexical and iconic signs have a relatively close match between the structures of the signifier and the signified. In symbols, however, the signifier and the signified are inherently discordant due to the arbitrary nature of the former. Importantly, the structural similarity of mental spaces is not an all-or-nothing property, allowing for an account of gradual emergence of symbols.

Furthermore, every sign includes an interpretant, i.e. the way it is interpreted by its receiver (Peirce, 1955). Because of this, communication should be analyzed relative to an animal's Umwelt, i.e. the way the world presents itself to it (Uexküll, 1992). In this context the non-arbitrariness of indexical and iconic signs as precursors to more complex blending in symbols becomes especially important. If we look closely at Umwelts of other animals, iconicity and indexicality can be found much more readily than before. Perhaps most clearly it can be seen in the often overlooked case of olfaction, as behaviors like marking an object with alarm odor, sexual pheromone, or urine marks can be interpreted as indexical due to their deictic function of demarcating presence, location, dominance, and territory. Additionally, the smells of those marks bear resemblance to the organism which produced it, suggesting an iconic relation as well (Parsons et al., 2018). A clear example is the urine marking of fire hydrants by dogs. McGuire, Olsen, Bemis, and Orantes (2018) suggest that small dogs lift their legs significantly higher than larger dogs for urine marking. The hydrant acts as a dominance display record, where the highest urine mark indexes the presence of a higher ranking individual. The suggestion is that dogs arriving at the fire hydrant interpret it as an iconic/indexical sign of their conspecifics in the area. The spatially arranged layout of the urine samples is interpreted as systematically mapping onto such properties of those individuals as size or dominance. Additionally, it might afford the perception of information about the temporal order of other dogs' visits to the hydrant or about their relative size. Such an understanding is not accessible to humans due the different Umwelt we inhabit (Quignon et al., 2003).

Turning to humans, language can be viewed as a high level hyper-blend (Turner, 2014), i.e. a blend using other blends as input mental spaces: in language the linguistic conceptual space is blended with a production/perception modality, most typically audition. Importantly, the semantic space of human language has a higher dimensionality than that of audition, which, leaving aside timing and timbre, is often estimated to be 2D (Keller, 2017). Recent experimental evidence suggests that iconicity is harder to maintain the more conflicting the topologies of the meaning and the signal space are, e.g. if the meaning space has higher dimensionality (Little, Eryılmaz, & De Boer, 2017). In such cases, more compositional or combinatorial structure is likely to emerge[1], leading to the development of more symbolic signs. This is indirectly supported by the observation that sign languages typically exhibit higher levels of iconicity than spoken languages (Taub, 2001). Interestingly, olfaction is often argued to have a more highly dimensional psychophysical space compared to other senses (Mamlouk & Martinetz, 2004; Magnasco, Keller, & Vosshall, 2015). While human olfactory abilities are often underestimated (Majid, 2021), humans exhibit abundant variation in olfactory receptor protein genes (Hasin-Brumshtein, Lancet, & Olender, 2009), which affect odor perception (Trimmer et al., 2019), and have many more nonfunctional olfactory receptor genes compared to other primates and other animals (Gilad, Man, Pääbo, & Lancet, 2003; Quignon et al., 2003). This could plausibly lead to more variability in Umwelt structure for smell compared to other senses between humans, and raises the possibility that olfaction – a high-dimensional perceptual modality affording more means for iconic communication – is less readily available for communication to us than to other animals.

Taken together, the arguments above could suggest that the uniquely high reliance on symbols in humans might be partially explained by the specifics of our perceptual and communicative Umwelt.

---

[1] This can be seen as a way to increase the dimensionality of the signal space: each compositional/combinatorial element serves as a new temporally separated dimension.

## References

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain.* WW Norton & Company.

Fauconnier, G., & Turner, M. (2008). *The way we think: Conceptual blending and the mind's hidden complexities.* Basic books.

Gilad, Y., Man, O., Pääbo, S., & Lancet, D. (2003). Human specific loss of olfactory receptor genes. *Proceedings of the National Academy of Sciences*, *100*(6), 3324–3327.

Hasin-Brumshtein, Y., Lancet, D., & Olender, T. (2009). Human olfaction: from genomic variation to phenotypic diversity. *Trends in Genetics*, *25*(4), 178–184.

Keller, A. (2017). *Philosophy of olfactory perception.* Springer.

Little, H., Eryılmaz, K., & De Boer, B. (2017). Signal dimensionality and the emergence of combinatorial structure. *Cognition*, *168*, 1–15.

Magnasco, M. O., Keller, A., & Vosshall, L. B. (2015). On the dimensionality of olfactory space. *bioRxiv*, 022103.

Majid, A. (2021). Human olfaction at the intersection of language, culture, and biology. *Trends in Cognitive Sciences*, *25*(2), 111–123.

Mamlouk, A. M., & Martinetz, T. (2004). On the dimensions of the olfactory perception space. *Neurocomputing*, *58*, 1019–1025.

McGuire, B., Olsen, B., Bemis, K., & Orantes, D. (2018). Urine marking in male domestic dogs: honest or dishonest? *Journal of Zoology*, *306*(3), 163–170.

Parsons, M. H., Apfelbach, R., Banks, P. B., Cameron, E. Z., Dickman, C. R., Frank, A. S., Jones, M. E., McGregor, I. S., McLean, S., et al.. (2018). Biologically meaningful scents: a framework for understanding predator–prey research across disciplines. *Biological Reviews*, *93*(1), 98–114.

Peirce, C. S. (1955). *Philosophical writings of peirce.* Dover Publications.

Quignon, P., Kirkness, E., Cadieu, E., Touleimat, N., Guyon, R., Renier, C., Hitte, C., André, C., Fraser, C., & Galibert, F. (2003). Comparison of the canine and human olfactory receptor gene repertoires. *Genome biology*, *4*(12), 1–9.

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in american sign language.* Cambridge University Press.

Trimmer, C., Keller, A., Murphy, N. R., Snyder, L. L., Willer, J. R., Nagai, M. H., Katsanis, N., Vosshall, L. B., Matsunami, H., & Mainland, J. D. (2019). Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proceedings of the National Academy of Sciences*, *116*(19), 9475–9480.

Turner, M. (2014). *The origin of ideas: Blending, creativity, and the human spark.* Oxford University Press.

Uexküll, J. von. (1992). A stroll through the worlds of animals and men: A picture book of invisible worlds.

# CALL COMBINATIONS IN CHIMPANZEES

MAËL LEROUX[*1,2,3], ANNE M. SCHEL[4], CLAUDIA WILKE[1,2,3], BOSCO CHANDIA[2], KLAUS ZUBERBÜHLER[2,3,5,6], KATIE E. SLOCOMBE[7] AND SIMON W. TOWNSEND[1,3,8]

[*]Corresponding Author: mael.leroux@uzh.ch
[1] Department of Comparative Linguistics, University of Zürich, Zürich, Switzerland
[2] Budongo Conservation Field Station, Masindi, Uganda
[3] Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zürich, Zürich, Switzerland
[4] Animal Behaviour and Cognition, Utrecht University, Utrecht, Netherlands
[5] Department of Comparative Cognition, University of Neuchâtel, Neuchâtel, Switzerland
[6] School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK
[7] Department of Psychology, University of York, York, UK
[8] Department of Psychology, University of Warwick, Warwick, UK

Through combining a finite set of words into larger compositional phrases, human language can express an open-ended, limitless number of messages. This *syntactic* capacity has been argued to be a key feature distinguishing language from any other non-human animal communication system. However, recent experimental evidence of syntactic-like structuring in monkeys has challenged this assumption and suggests syntax might be evolutionary more ancient with its origins deeply rooted in the primate lineage. Comparable data in great apes, our closest-living relatives, are central to reconstructing the more recent evolutionary history of syntax, yet are currently lacking. In this study, we address this issue and provide the first robust evidence for rudimentary syntactic structuring in our closest-living relatives, chimpanzees (though see Crockford and Boesch 2005; Leroux et al. 2021, for evidence of combinatoriality). Chimpanzees produce "*alarm-hoos*" when surprised or frightened and "*waa-barks*" when potentially recruiting other group members during aggression or hunting. Anecdotal data suggested chimpanzees also combine these two calls together, specifically when encountering a snake while being isolated from the main group. Through presenting wild chimpanzees with model snakes we i) confirmed combinations were more likely to occur when an individual encountered the snake separated from the group and ii) investigated receivers' responses showing they were more likely to join the caller after hearing the call combination compared with singly-occurring "*alarm-hoos*". Finally, to confirm these findings and verify the meaning-bearing nature of the call combination, we conducted systematic experiments, playing back an artificial call combination and both calls produced independently. Chimpanzees reacted most strongly to the combination, showing quicker and longer responses, compared with both individual calls. We conclude the "alarm-hoo + waa-bark" combination represents a compositional syntactic-like structure, where the meaning of the sequence is derived from the meaning of its parts. Our work supports previous work in monkeys and indicates language is unlikely to have evolved de novo in the human lineage but rather the cognitive building blocks facilitating syntax were already present in our last common ancestor with chimpanzees and are perhaps even older.

## Acknowledgements

We are grateful to the staff of the Budongo Conservation Field Station for supporting us with this study, and to the Royal Zoological Society of Scotland for providing core funding. We also thank the Ugandan Wildlife Authority, the Ugandan National Council for Science and Technology and the President's Office for permission to conduct this research. We thank Denis Lomoro for helping with data collection.

## References

Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, *142*, 397–421.

Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant-hoots with food-calls into larger structures. *Animal Behaviour*, *179*, 41–50. https://doi.org/10.1016/j.anbehav.2021.06.026

# COMMUNITY SIZE INFLUENCES WORD ORDER

SHIRI LEV-ARI 1[*1,]

[*]Shiri.Lev-Ari@rhul.ac.uk
[1]Psychology Department, Royal Holloway, University of London, Egham, UK

There are six possible ways to order Subject, Verb, and Object in a sentence, but the distribution of these orders across the world's languages is highly skewed: SVO and SOV are more than 100 times more common than OSV (Dryer, 2005). It is often hypothesized that the preferred word orders reflect cognitive biases in event perception and optimization for communication (e.g., Gibson et al., 2013; Hahn et al., 2020; Schouwstra & de Swart, 2014). Then why do not all languages use these orders? This study tests whether languages are more likely to manifest universal tendencies if they face more communicative challenges. It specifically tests whether larger communities are more likely to use preferred word orders as a way to overcome the greater communicative challenges they encounter.

Larger communities face greater communicative challenges: they have greater input variability, less shared history, and information takes longer to travel. Larger communities overcome these challenges by creating more robust languages, including languages with more systematic grammar (Lupyan & Dale, 2010; Raviv et al., 2019) and greater sound symbolism (Lev-Ari et al., 2021). These languages thus align with prior cognitive biases (e.g., sound-meaning associations) to facilitate communication. Reliance on word orders that fit cognitive biases can similarly facilitate communication.

To test this hypothesis, 15 groups of 8 participants (large groups) and 15 groups of 4 participants (small groups) played a communication game in dyads for 14 rounds, changing partner every round. Participants described videos from Hall et al. (2013) to their partner by ordering pictures of the subject, object, and action. The partner guessed which video they saw and both received feedback. Then they switched roles. Thirty additional participants played the game alone (solo).

To test whether group size influences reliance on universally favored word orders, word order on each trial was coded as 1 if it was SVO or SOV, or 0 otherwise. A logistic mixed effects model was conducted with Group Size (solo, small, large; ref level=large), Round (scaled), their interaction, as well as Participant's Native Word Order as fixed effects, and Participant nested within Session, Session, and Items as random variables. It revealed that reliance on a common word order increased with time for large groups (β=0.57, p<0.001) but that it did not increase as much among the small groups (β=-0.18, p<0.01) and solo individuals (β=-0.24, p<0.001). The plot suggests that this is driven by SVO order dominance. These results indicate that common word orders, especially SVO, are driven by communicative pressures. The fact that the use of common word orders increases with round, that it is significant even after controlling for the word order in participants' native language, and that it depends on group size, indicates that performance was not driven by knowledge of native language but by the properties of the interaction.

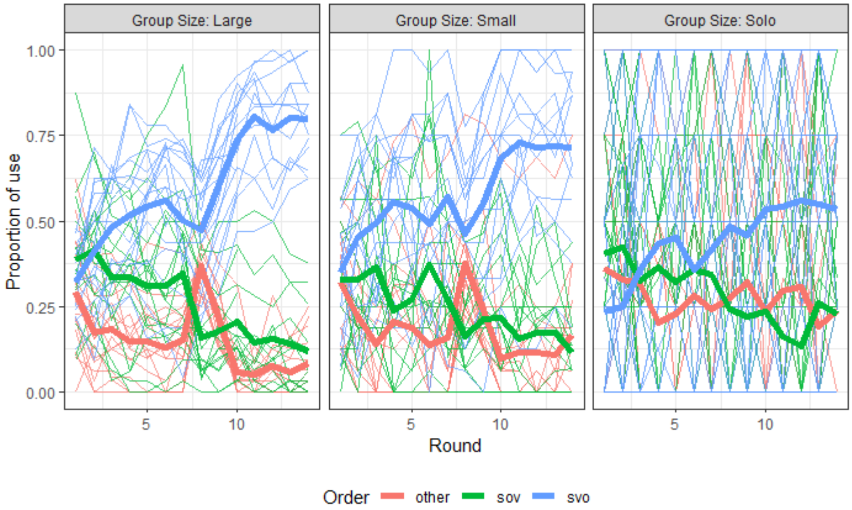


Figure 1. Word Order as dependent on Round and Group Size. Thin lines show each session, thick lines show condition averages.

The results of the study suggest that cross-linguistic variability in word order might reflect differences in communicative pressures, with less common word orders being more common in smaller communities. A preliminary survey of 1047 languages from WALS (Haspelmath et al., 2005) crossed with population data from Ethnologue (Eberhard et al., 2020) supports this prediction, especially for SVO order vs others (in line with experimental results; see figure 1). Future research should further examine how cross-cultural differences in communicative pressure influence exhibition of universally-favored patterns.

## Acknowledgements

This research was supported by the BA/Leverhulme Small Research Grants scheme (SRG1920\101183).

## References

Dryer, M. S. (2005). Order of subject, object and verb. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures* (pp. 330–333). Oxford, England: Oxford University Press.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2020). *Ethnologue: Languages of the world* (23rd ed.). Dallas, TX: SIL International

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological science*, *24*(7), 1079-1088.

Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, *117*(5), 2347-2353.

Hall, M. L., Mayberry, R. I., & Ferreira, V. S. (2013). Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition*, *129*(1), 1-17.

Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (2005). *The world atlas of language structures*. OUP Oxford.

Lev-Ari, S., Kancheva, I., Marston, L., Morris, H., Swingler, T., & Zaynudinova, M. (2021). “Big” Sounds Bigger in More Widely Spoken Languages. *Cognitive Science*, *45*(11), e13059.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1), e8559.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*(1907), 20191262.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, *131*(3), 431-436.

# LANGUAGE ACQUISITION AND LANGUAGE EVOLUTION: LESSONS FROM DUAL INHERITANCE THEORY

Josh Linden

Correspondence: jsl169@georgetown.edu
Department of Linguistics, Georgetown University, Washington, DC, United States

Dual Inheritance Theory (DIT; Richerson & Boyd 1978; Russell & Muthukrishna 2021) is emerging in biological anthropology and related fields as a leading candidate for a plausible and highly explanatory account of the recent evolutionary history of our own species. A growing canon of literature spanning many decades has explored its implications in detail and found robust empirical support across the social sciences. The theory is built on three core assertions: 1) that early hominins gained the genetic capacity to acquire and transmit information culturally as well as genetically, 2) that such cultural information would itself be subject to natural selection, giving rise to a second system of inheritance operating on a distinct set of selectional principles and 3) that *culture-gene coevolution* would allow the human genome to continue to optimize for the changing conditions created by the cultural environment (Henrich & McElreath 2007). Together with the logic of evolution by natural selection, these axioms give rise to a detailed outline of the development of human cognition with consistent empirical support across such diverse domains as evolutionary biology (Herrmann et al. 2007), evolutionary psychology (Chudek & Henrich 2015), social psychology (Gervais et al. 2021), psychoanalysis (Whitebook 2019), psychological anthropology (Paul 2018), anthropology (Henrich 2016), sociobiology (Boyd & Richerson 1980), economics (Laing 2008), animal cognition (Horner & Whiten 2004) and many others.

Nonetheless, DIT remains all but absent from theoretical linguistics and its vast implications for the study of language and particularly of language evolution have not been explored. There has been relatively little agreement among theoretical linguists about even how to study language evolution, although some significant progress has begun to appear in the last decade. Fitch (2017) insightfully suggests a multicomponent approach, in which we refrain from assuming language was *an* innovation and instead ask which innovations might make it up, and he introduces the term *Derived Component of Language (DCL)* to refer to any element of language that we hypothesize arose in humans or our immediate ancestors after our evolutionary divergence from chimpanzees. He further points out the necessity of creating multiple plausible hypotheses and comparing them to each other experimentally rather than considering each model's plausibility in isolation.

Adopting and extending this framework, the current inquiry explores the theoretical implications of assuming that language evolution happened primarily as described by DIT: as a series of culturally transmitted innovations achieved on human timescales and continuously adapted to by our genetics on longer timescales. I show how this assumption reinforces Fitch's (2017) *staged-protolanguages model* approach, and I create an explicitly nonspecific Fitchian model in which the uniquely human elements of language arise in some unknown order $\mathrm{DCL}_1$, $\mathrm{DCL}_2$, …, $\mathrm{DCL}_n$. Situating these as a series of coevolved traits arising within DIT, I extrapolate the model into the present and conclude that the empirical facts of language acquisition are especially likely to differ under competing Fitchian hypotheses.

Informed by existing DIT models of how information is transmitted between communities and across generations (e.g. Boyer 1998; Linquist 2007), I propose that two domains featuring partial acquisition are of most pressing interest to the evolutionary linguist: the overlapping but distinct stages of linguistic competence we observe in first-language (L1) learners, and the empirical facts of *homesign*. The former include the babbling stage, first words, the two-word stage, the multi-word stage and an extended process of refinement and complexification (Salim & Mehawesh 2014). The latter refers to the structured communicative systems that arise in Deaf children growing up without exposure to a signed language (Franklin et al. 2011; Coppola & Newport 2005). Homesigners reliably engage in manual babbling, invent signs with diverse meanings and begin to combine them to enter the two-word and multi-word stages (Morford 1996). They do not, however, develop all further elements of Universal Grammar, and this is highly informative in light of the coevolutionary optimization loops the present model describes. I borrow the term *cognitive dependency* (Seuren 2006) from psycholinguistics and cognitive semantics, which I formalize within the DIT framework by analogy to *exaptation* (Gould & Vrba 1982) in evolutionary biology. I then use it to argue that if $\mathrm{DCL}_i$ is reliably observed in homesign or at some stage of L1 acquisition while $\mathrm{DCL}_j$ is not, it is likely that $i < j$. Finally I consider briefly which potential orderings of DCLs preliminarily appear best supported.

## References

Boyd, Robert & Peter J. Richerson. 1980. Sociobiology, culture and economic theory. *Journal of Economic Behavior & Organization* 1(2). 97–121. doi:10.1016/0167-2681(80)90005-0.

Boyer, Pascal. 1998. Cognitive tracks of cultural inheritance: How evolved intuitive ontology governs cultural transmission. *American Anthropologist* 100(4). 876–889. doi:10.1525/aa.1998.100.4.876.

Chudek, Muthukrishna M & J Henrich. 2016. Cultural evolution. Essay. In , *The Handbook of Evolutionary Psychology*. Hoboken, NJ: John Wiley & Sons, Inc.

Coppola, M. & E. L. Newport. 2005. Grammatical subjects in home sign: Abstract linguistic structure in adult primary gesture systems without linguistic input. *Proceedings of the National Academy of Sciences* 102(52). 19249–19253. doi:10.1073/pnas.0509306102.

Fitch, W. Tecumseh. 2017. Empirical approaches to the study of Language Evolution. *Psychonomic Bulletin & Review* 24(1). 3–33. doi:10.3758/s13423-017-1236-5.

Franklin, Amy, Anastasia Giannakidou & Susan Goldin-Meadow. 2011. Negation, questions, and structure building in a homesign system. *Cognition* 118(3). 398–416. doi:10.1016/j.cognition.2010.08.017.

Gervais, Will M., Maxine B. Najle & Nava Caluori. 2021. The origins of religious disbelief: A dual inheritance approach. *Social Psychological and Personality Science* 12(7). 1369–1379. doi:10.1177/1948550621994001.

Gleitman, Lila R. & Elissa L. Newport. 1995. The Invention of Language by Children: Environmental and Biological Influences on the Acquisition of Language. Essay. In , *Language*, 1–24. Cambridge, MA: MIT Press.

Gould, Stephen Jay & Elisabeth S. Vrba. 1982. Exaptation—a missing term in the science of form. *Paleobiology* 8(1). 4–15. doi:10.1017/s0094837300004310.

Henrich, Joseph & Richard McElreath. 2007. Dual-inheritance theory: the evolution of human cultural capacities and cultural evolution. Essay. In , *Oxford Handbook of Evolutionary Psychology*. Oxford: Oxford University Press.

Henrich, Joseph Patrick. 2016. *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton: Princeton University Press.

Herrmann, Esther, Josep Call, Hernàndez-Lloreda María Victoria, Brian Hare & Michael Tomasello. 2007. Humans have evolved specialized skills of social cognition: The Cultural Intelligence Hypothesis. *Science* 317(5843). 1360–1366. doi:10.1126/science.1146282.

Horner, Victoria & Andrew Whiten. 2004. Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (homo sapiens). *Animal Cognition* 8(3). 164–181. doi:10.1007/s10071-004-0239-6.

Laing, Bart Du. 2008. Dual inheritance theory, contract law, and institutional change – towards the co-evolution of behavior and institutions*. *German Law Journal* 9(4). 491–514. doi:10.1017/s2071832200006556.

Linquist, Stefan. 2007. Prospects for a dual inheritance model of emotional evolution. *Philosophy of Science* 74(5). 848–859. doi:10.1086/525626.

Morford, Jill P. 1996. Insights to language from the study of gesture: A review of research on the gestural communication of non-signing deaf people. *Language & Communication* 16(2). 165–178. doi:10.1016/0271-5309(96)00008-0.

Paul, Robert A. 2018. Culture from the perspective of dual inheritance. *Advances in Culture Theory from Psychological Anthropology*. 47–73. doi:10.1007/978-3-319-93674-1_3.

Richerson, P & R Boyd. 1978. A dual inheritance model of the human evolutionary process I: Basic postulates and a simple model. *Journal of Social and Biological Systems* 1(2). 127–154. doi:10.1016/s0140-1750(78)80002-5.

Richie, Russell, Charles Yang & Marie Coppola. 2014. Modeling the emergence of Lexicons in Homesign Systems. *Topics in Cognitive Science* 6(1). 183–195. doi:10.1111/tops.12076.

Russell, Connair J. & Michael Muthukrishna. 2021. Dual inheritance theory. *Encyclopedia of Evolutionary Psychological Science*. 2140–2146. doi:10.1007/978-3-319-19650-3_1381.

Salim, Jamal Azmi & Momammad Mehawesh. 2014. Stages in language acquisition: A case study. *English Language and Literature Studies* 4(4). doi:10.5539/ells.v4n4p16.

Seuren, P.A.M. 2006. Meaning: Cognitive dependency of lexical meaning. *Encyclopedia of Language & Linguistics*. 575–577. doi:10.1016/b0-08-044854-2/01056-7.

Whitebook, Joel. 2019. Our dual inheritance: On psychoanalytic social theory Today. *American Imago* 76(1). 1–14. doi:10.1353/aim.2019.0000.

# SYNONYMS AND HOMONYMS IN SIGNALING GAME – WHICH ARE STABLE AND WHEN?

DOROTA LIPOWSKA*[1]

*Corresponding Author: lipowska@amu.edu.pl
[1]Chair of Linguistic Methodology, Adam Mickiewicz University in Poznań, Poland

Evolution and structure of language is often analysed using computational modelling (Cangelosi & Parisi, 2002; Nolfi & Mirolli, 2010; D'Ulizia, Ferri, & Grifoni, 2020). A particularly appealing research paradigm is inspired by the idea that language might have spontaneously appeared in a population of communicating individuals, possibly with some adaptive features (Pinker & Bloom, 1990). This standpoint prompted numerous analysis of multi-agent models, which mimic such communication and try to infer the properties of the emerging language and its possible further evolution (Steels, 2012; Gong, Shuai, & Zhang, 2014; Kirby, Griffiths, & Smith, 2014).

In certain models of this kind, language emergence and evolution is studied using the signaling game (Lewis, 2002), where communicating agents must decide which signal to send or how to interpret the signal they have received. To cope with this, agents very often use some form of the reinforcement learning (Skyrms, 2010; Lenaerts, Jansen, Tuyls, & De Vylder, 2005; Barrett, 2006; Franke, 2016; Mühlenbernd & Franke, 2012; Lipowska & Lipowski, 2018).

By taking signals as words, we can regard the form-meaning mapping that emerges during the signaling game as language. The mapping can be a one-to-one correspondence (in a signaling game terminology, it is a signaling system), but there are also other possibilities, namely, homonyms or synonyms can emerge. Neglecting some linguistic nuances (Ravin & Leacock, 2000), we can posit that in the case of homonymy, one word carries different meanings while synonymy means that a single concept is expressed by different words.

It seems quite plausible that synonyms or homonyms change in time. For example, the frequency of their usage may change and gradually one form will be preferred over the other, and the latter eventually can even disappear. Although linguistic data are difficult to interpret, there are some indications that in natural languages synonyms are quite rare in contrast to homonyms, which appear to be more common (Hurford, 2003; Clark, 1990). Some linguists even insist that true synonyms do not exist or at best are very rare compared to homonyms (Lyons, 1981; Goldberg, 1995). There are some arguments that the difference in the frequency of synonyms and homonyms may be due to evolutionary pressures favouring speakers rather than hearers (Hurford, 2003), or to language acquisition in childhood (Markman, 1989).

Let us notice that synonyms actually compete in a quite different way from homonyms, which can be demonstrated already within the framework of the signaling game. While synonymous words compete for being selected by a speaker, for a homonymous word, it is the hearer's role to assign an appropriate interpretation. It is thus possible that such a difference can affect an overall dynamics of synonyms and homonyms and eventually result in different degrees of their prevalence.

We approached the problem of evolution and stability of synonyms and homonyms using computational modelling. Within the framework of the signaling game, we show that the reinforcement learning should operate in the so-called super-linear regime with probabilities of selections increasing faster than linearly with the accumulated weights. The linear regime would, instead, lead to languages with very stable synonyms and relatively fast decaying homonyms, which is probably at odds with some linguistic observations (Hurford, 2003). Our work indicates that the prevalence and evolution of synonoyms and homonyms in natural languages may give us some valuable clues as to the nature of the mechanisms that drive linguistic processes. More details concerning our model and the results obtained can be found in (Lipowska & Lipowski, 2022).

Perhaps an interesting question is why nonlinear rather than the simplest (maybe naively expected) linear feedback drives linguistic processes. Related studies, under mathematically similar setup of the so-called urn models, in certain marketing or economic contexts showed that the value of the system grows faster than linearly with the number of users. Such a behaviour seems to characterize competition between, e.g., video formats, operating systems, and even types of keyboards. In the literature, it is referred as Metcalfe's Law (Shapiro & Varian, 1998; Arthur, 1994). In the signaling game and in linguistic context, it would mean that a benefit of using a certain word (and thus a probability of its future selection) increases faster than linearly with the number of successful communications. Considering the complexities of language evolution, with its various social, biological, and cognitive aspects, it seems quite likely.

## References

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy.* University of Michigan Press.

Barrett, J. A. (2006). *Numerical simulations of the Lewis signaling game: Learning strategies, pooling equilibria, and the evolution of grammar.* UC Irvine: Institute for Mathematical Behavioral Sciences.

Cangelosi, A., & Parisi, D. (Eds.). (2002). *Simulating the evolution of language.* London: Springer-Verlag.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*(02), 417–431.

D'Ulizia, A., Ferri, F., & Grifoni, P. (2020). A survey on modeling language evolution in the new millennium. *New Generation Computing*, *38*(1), 97–124.

Franke, M. (2016). The evolution of compositionality in signaling games. *Journal of Logic, Language and Information*, *25*(3), 355–377.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Gong, T., Shuai, L., & Zhang, M. (2014). Modelling language evolution: Examples and predictions. *Physics of Life Reviews*, *11*(2), 280–302.

Hurford, J. R. (2003). Why synonymy is rare: Fitness is in the speaker. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in Artificial Life. Proceedings of the Seventh European Conference on Artificial Life ECAL 2003* (Vol. 2801, pp. 442–451). Berlin, Heidelberg: Springer-Verlag.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114.

Lenaerts, T., Jansen, B., Tuyls, K., & De Vylder, B. (2005). The evolutionary language game: An orthogonal approach. *Journal of Theoretical Biology*, *235*(4), 566–582.

Lewis, D. K. (2002). *Convention: A philosophical study.* Oxford, UK: Blackwell.

Lipowska, D., & Lipowski, A. (2018). Emergence of linguistic conventions in multi-agent reinforcement learning. *PLoS ONE*, *13*(11), e0208095.

Lipowska, D., & Lipowski, A. (2022). Stability and evolution of synonyms and homonyms in signaling game. *Entropy*, *24*(2), 194.

Lyons, J. (1981). *Language, meaning and context.* London: Fontana.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* MIT Press.

Mühlenbernd, R., & Franke, M. (2012). Simulating the emergence of conventions in small-world networks. In *Proceedings of the 21st International Conference on Behavior Representation in Modeling & Simulation (BRIMS 2012)* (pp. 37–42). Red Hook, NY: Curran Associates, Inc.

Nolfi, S., & Mirolli, M. (Eds.). (2010). *Evolution of Communication and Language in Embodied Agents.* Berlin, Heidelberg: Springer-Verlag.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–727.

Ravin, Y., & Leacock, C. (2000). *Polysemy: Theoretical and Computational Approaches.* Oxford, UK: Oxford University Press.

Shapiro, C., & Varian, H. R. (1998). *Information Rules: A Strategic Guide to the Network Economy.* Harvard Business Press.

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information.* Oxford, UK:

Oxford University Press.

Steels, L. (Ed.). (2012). *Experiments in Cultural Language Evolution* (Vol. 3). Amsterdam: John Benjamins.

# HOW MUCH DO WE AGREE ON WHAT WORDS MEAN?

GARY LUPYAN*[1], STEVEN PIANTADOSI[2], LOUIS MARTÍ[2], CELESTE KIDD[2],
ELLA LIU[1], JEROEN VAN PARIDON[1] and YANCHAO BI[3]

*Corresponding Author: lupyan@wisc.edu
[1]Dept of Psychology, University of Wisconsin-Madison, USA [2]Dept of Psychology, UC Berkeley, USA [3]IDG/McGovern Institute for Brain Research, Beijing Normal University, China

At the core of every language is a vocabulary—a set of building blocks from which language users can construct arbitrarily complex meanings. Vocabularies are, of course, learned. Members of a speech community must learn the form of each word—its phonology—but also, critically, its meaning. Where do word meanings come from? And how similar are the meanings of the same word in different people within a speech community? These questions are of central importance to the study of language evolution for at least three reasons. First, understanding the emergence of vocabularies requires that we understand the extent to which words demarcate cognitively privileged categories (i.e., words label our concepts, e.g., Snedeker & Gleitman, 2004), as opposed to demarcating categories that are shaped by communicative needs and history—categories that might not be learned in the absence of labels (words help create concepts, e.g., Lupyan et al., 2007; Thompson et al., 2020). Second, for languages to function as effective communicative systems, it is generally thought that people must closely agree on what words mean (e.g., Hutchins & Hazlehurst, 2006). If word meaning variability within a speech community is pervasive, it would raise the question of how much agreement is really necessary for effective communication. Third, if word meaning variability is indeed pervasive, how can we use language to coordinate as effectively as we seem to? When we talk, misunderstandings seem to be more the exception than the norm. Here, we bring together several sources of recent evidence collected in our respective labs and observed in published word norms that suggest that within a speech community, differences in meanings for everyday words may be more extensive than previously realized (cf. Clark, 1998).

There is no gold standard for quantifying and comparing word meanings, and so we sought to estimate variability in a few different ways: (1) Sorting, e.g., sort these words according to how similar they are to one another. (2) Similarity judgments, e.g., is a penguin more similar to a whale or a seal? (3) Elicitation and endorsement of word meanings, e.g., what does “energy” mean? Which of these meanings of “energy” best approximates your meaning? (4) Comparison of people’s judgments of semantic dimensions such as concreteness. (5) Beyond explicit behaviors, we can also compare (using fMRI) neural representations elicited by the same word in different people.

Each measure revealed clear similarities, e.g., people clustered body parts together and separately from animals, clustered positive emotions together and separately from negative emotions. In general, agreement for concrete words was greater than for abstract words. But substantial variability existed even for concrete words. Fig. 1A shows an example of how two people sorted a group of common animals. Sorting-based correlations of such concrete words rarely exceed *r*=.5 (Wang & Bi, 2021). We also saw substantial variability when using similarity probes, e.g., when asked whether a seal is more similar to a penguin or a whale, 44% of people chose “whale”. Interestingly, people were largely unaware that judgments like these produce divergent responses, believing a large majority will respond as they themselves did. Those who responded with “whale” thought that 75% would do the same (Martí et al., 2021). Fig. 1B shows endorsement patterns of various senses of “energy” (generated by a separate set of participants), revealing three distinct profiles. Fig. 1C shows concreteness ratings (Brysbaert et al., 2014). While many words show expected unimodal distributions, many others show clear bimodality hinting at systematic differences in how different people construe these words.

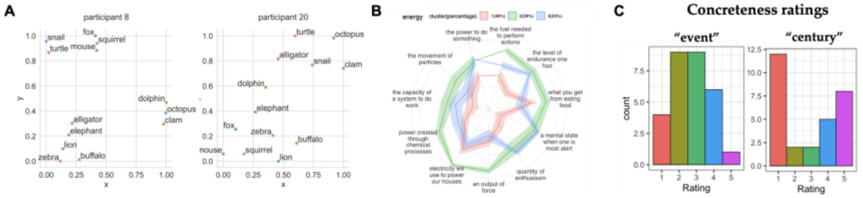


Figure 1. (A) Two solutions on an animal-sorting task. (B) Results from a meaning elicitation and endorsement task for the word “energy”. (C) Distributions of concreteness ratings for two example words (rating of 1 corresponds to maximally abstract).

Do results such as these indicate true variability? If so, where do these differences come from? We consider several possibilities including different learning biases, different sensorimotor experiences, and different linguistic experiences. How do people communicate in the presence of these differences? We will discuss three possibilities: (1) These differences have no consequences for everyday communication, only showing up in specific contexts. (2) Misalignments are quickly repaired (Healey et al., 2018) or compensated through pragmatic inference. (3) Consequential errors in communication are more common than generally acknowledged.

## References

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Clark, H. H. (1998). Communal Lexicons. In K. Malmkjær & J. Williams (Eds.), *Context in Language Learning and Language Understanding* (p. 63). Cambridge University Press.

Healey, P. G. T., Mills, G. J., Eshghi, A., & Howes, C. (2018). Running Repairs: Coordinating Meaning in Dialogue. *Topics in Cognitive Science*, *10*(2), 367–388. https://doi.org/10.1111/tops.12336

Hutchins, E., & Hazlehurst, B. (2006). *How to invent a lexicon: The development of shared symbols in interaction*. https://doi.org/10.4324/9780203993699-17

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1082.

Martí, L., Wu, S., Piantadosi, S. T., & Kidd, C. (2021). *Latent diversity in human concepts*. PsyArXiv. https://doi.org/10.31234/osf.io/fb8s2

Snedeker, J., & Gleitman, L. (2004). Why is it hard to label our concepts? In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon* (illustrated edition, pp. 257–294). The MIT Press.

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*, 1–10. https://doi.org/10.1038/s41562-020-0924-8

Wang, X., & Bi, Y. (2021). Idiosyncratic Tower of Babel: Individual Differences in Word-Meaning Representation Increase as Word Abstractness Increases. *Psychological Science*, 09567976211003877. https://doi.org/10.1177/09567976211003877

# PHONOLOGICAL REPRESENTATION IN BONOBOS *(PAN PANISCUS)* AS REVEALED THROUGH VOCABULARY ERRORS

HEIDI LYN[*1] and MARK YATES[2]

[*]Corresponding Author: hlyn@southalabama.edu
[1]Joan M. Sinnott Chair of Psychology and Department of Marine Sciences, University of South Alabama, Mobile AL USA
[2]Department of Psychology, University of South Alabama, Mobile AL USA

## 1. Objective

Human minds contain a complex set of associations with each word, morpheme, and phoneme in our language. When we mis-speak, those associations make some errors more likely than others (Fromkin, 1980). This study seeks to clarify the possible phonological associations within the mind of language-using bonobos.

## 2. Methods

Participants were two bonobos *(Pan paniscus)* that had been reared in a language-enriched environment and shown to understand English sentences similar to the abilities of a 2.5 year old child (Savage-Rumbaugh et al., 1993). Over the course of 10 years, these apes were given vocabulary tests regularly and all responses made to the lexigram keyboard were recorded and coded for a previous study (Lyn, 2007). These errors were shown to be non-random and associated with visual, conceptual, and auditory representations of the item. For example, lemons-lemonade, orange-apple, and cereal-milk.

To further explore the auditory components of these representations, phonological transcriptions of the errors were coded as a match/no match from the sample to the error. All examples that included shared morphemes (e.g. Orange-orange juice) were removed, resulting in 1400 example of sample and error pairs.

## 3. Results

The bonobos made significantly more errors that matched the sample in phoneme than would be expected by chance (see Fig 1). Their responses varied by vowel and consonant, with more consonant matches in the first and third positions, and more vowel matches in the second and fourth positions. For example, BIG and BUG match in the first and third phoneme (bIg and b^g) and JUICE and FOOD match in the second phoneme (jus and fud).

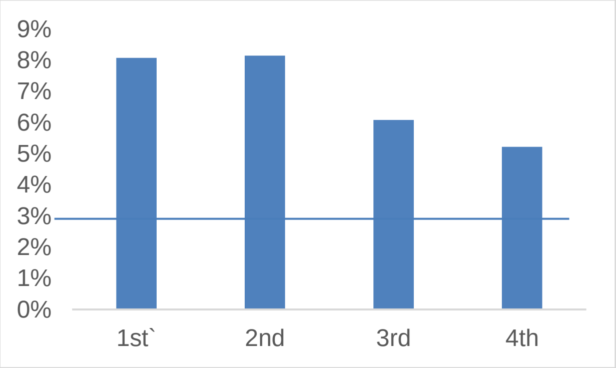


Figure 1. Percentage of phonological matches across the first four positions of sample and error. Line represents conservative estimate of chance, 1/26 (the number of phonemes in each position of the samples and errors ranged from 26-32). All are p<.001, binomial test.

## 4. Discussion

Our results suggest that the web of associations in the mind of language using bonobos is more elaborate than simple auditory similarity and may include phonemic information. These findings further suggest that this building block of human language was present before the evolutionary split of humans and bonobos.

## Acknowledgements

Heidi Lyn holds the Joan M. Sinnott Chair which is funded by the University of South Alabama Foundation.

## References

Fromkin, V. A. (1980) *Errors in linguistic performance. Slips of the tongue, ear, pen and hand.* Academic Press, New York.

Lyn, H. (2007). Mental representation of symbols as revealed by vocabulary errors in two bonobos (Pan paniscus). *Animal Cognition, 10*(4), 461-475.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., & Bates, E. (1993). Language comprehension in ape and child. *Monographs of the society for research in child development*, i-252.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. Behavior Research Methods, Instruments, & Computers, 36(3), 481–487. https://doi.org/10.3758/BF03195594

# TESTING THE ‘ROUNDNESS’ AND ‘SHARPNESS’ OF NONWORDS AND SHAPES IN GREAT APES.

Konstantina Margiotoudi*[1] and Friedemann Pulvermüller [2,3,4]

*Corresponding Author: konstantina.margiotoudi@univ.amu.fr
[1]Laboratoire de Psychologie Cognitive, Aix-Marseille University, CNRS, Marseille, France
[2]Brain Language Laboratory, Department of Philosophy and Humanities, Freie Universität, Berlin, Germany
[3]Berlin School of Mind and Brain, Humboldt Universität, Berlin, Germany
[4]Einstein Center of Neurosciences, Berlin, Germany

Although sound symbolic effects of ‘bouba-kiki’ type have been reported across cultures (Ćwiek et al., 2022), it remains unclear whether this effect is an ability unique to humans. Sound symbolic effects between speech sounds and shapes (Köhler, 1919) might have played an important role in shaping protolanguages. Humans could have understood that certain speech sounds can evoke certain sensory properties. This ability could have plausibly assisted referential insight in our human ancestors (Perniss and Vigliocco, 2014). The discussion on the role of sound symbolism in shaping human languages gives rise to the following question: “Can nonhuman great apes detect congruencies between ‘round’ and ‘sharp’ nonwords and curved and sharp shapes?”. Previous research has shown that a group of touchscreen trained chimpanzees (N=6) and gorillas (N=2), when tested with a two-alternative forced choice task on sound-shape mappings showed no sound symbolic effects (Margiotoudi et al., 2019). Specifically, both gorillas and chimpanzees failed to detect that a ‘round’ nonword, such as “bouba”, is a good fit to a curved shape and a ‘sharp’ nonword, such as “kiki”, is a good fit to a spiky shape. In parallel, humans were tested with the same task and showed a significant sound symbolic congruency. In both tasks, neither the nonhuman nor the human great apes were explicitly instructed to detect congruencies between nonwords and shapes. Notably, when a second group of human participants were explicitly instructed to detect sound symbolic congruencies in the same task, they showed above chance performance and detected congruencies 10% more often compared to the previous implicit task. Conclusively, when humans are explicitly instructed to detect sound symbolic congruencies, they achieve a better performance. However, humans are exposed to speech stimuli compared to nonhuman great apes and have referential insight. Could a nonhuman great ape exposed to speech stimuli detect speech sound-shape consistencies? In a different study, a language competent bonobo was tested with a match-to-sample task on the 'bouba-kiki' effect. The bonobo has been able to match English words to pictures with the same task. Hence, he tries to pick the best match from different stimuli. Using his strategy, a second study investigated whether a nonhuman great ape being exposed to speech stimuli would be able to detect sound symbolic congruencies under an explicit match-to-sample task. Specifically, the bonobo performed his familiar match English words to pictures task, while sound symbolic trials of "bouba-kiki" type were interspersed. This kind of paradigm allowed testing a nonhuman primate under explicit instructions. The bonobo tried to guess the best matching shape for the 'round' and 'sharp' nonwords as he did for the English word-picture matching. The results of the study revealed no sound symbolic congruency detection but significantly above chance performance in matching English words to pictures (Margiotoudi et al., 2022).

These findings suggest that the 'bouba-kiki' mapping is plausibly an ability unique to humans. These results might be explained by neurobiological differences found between human and nonhuman great apes that are relevant to the mechanism supporting the speech sounds-shape mappings. We suggest that the mechanism of this effect is found in the knowledge of the multimodal products of our hand actions. This binding has special preconditions found at a neuronal level, with auditory-visual and motor systems linked in the brain via long white matter tracts.

## References

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., ... & Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200390.

Köhler W. 1929 Gestalt psychology. New York, NY: Liveright.

Margiotoudi, K., Allritz, M., Bohn, M., & Pulvermüller, F. (2019). Sound symbolic congruency detection in humans but not in great apes. *Scientific reports*, *9*(1), 1-12.

Margiotoudi, K., Bohn, M., Schwob, N., Taglialatela, J., Pulvermüller, F., Epping, A., ... & Allritz, M. (2022). Bo-NO-bouba-kiki: picture-word mapping but no spontaneous sound symbolic speech-shape mapping in a language trained bonobo. *Proceedings of the Royal Society B, 289*(1968), 20211717.

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in psychology, 1*, 227.

# VOCAL LEARNING IS AT ODDS WITH NON-VOCAL DEXTERITY

Pedro Tiago Martins*[1]

*Corresponding Author: pt.martins@ub.edu
[1]Psycholinguistics and Neurolinguistics Lab, University of Ljubljana, Ljubljana, Slovenia

## 1. Introduction

In this work, I propose that vocal learning complexity is at odds with non-vocal dexterity. In order words, given some conditions generically required for vocal learning behavior, species lacking appendages fine-tuned for grasping, object manipulation, etc. ('prehensile appendages') are more likely to display complex vocal behavior than species which do have and expertly use such appendages. I hypothesize that the use of the mouth or analogous structures (and those adjacent to it) for all or most object manipulation and interaction promotes complex vocal behavior. Conversely, the existence of appendages tuned for fine control relegates the vocal apparatus to simpler vocal, feeding, and manipulation behaviors. In conjunction (and indeed in constant interaction) with several other factors, this relationship helps shape the vocal learning phenotype.

This hypothesis has implications for how vocal learning and related abilities evolved across species, and emphasizes the importance of ecological and behavioral factors for this capacity. This is in contrast with much of the vocal learning literature which tends to be focused on neural mechanisms alone.

## 2. A closer look

The species to look at in order to study this relationship are *accepted* and *possible* vocal learners. Here I consider vocal learning in a broad sense (Ghazanfar, Liao, & Takahashi, 2019; Martins & Boeckx, 2020; Fischer, Wegdell, Trede, Dal Pesco, & Hammerschmidt, 2020), not tied to a very specific neural circuit nor limited to imitative behavior, which is crucial for the current hypothesis.

Given a set of factors in place (sound production anatomy, some kind of neural circuitry subserving control of that anatomy, functional and ecological pressures promoting and shaping vocal behavior), this hypothesis adds that well-developed prehensility in appendages not related to vocal behavior (e.g., "hands"), pushes species away from complex vocal behavior, while the lack of such appendages or dexterity thereof brings species closer to it. Indeed, for the most part, vocal learning is displayed most clearly by species with not a lot of dexterity (several bird orders, pinnipeds—flyers and swimmers), while other, more dexterous species with apparently the right mix of factors in place display it in a more arduous, peculiar, or simple manner—to the point where they are often not accepted as vocal learners (e.g., non-human primates, rodents).

The following independent reasons for this relationship are put forward as part of the hypothesis:

1. lack of manual dexterity puts emphasis on mouth and adjacent structures for object manipulation, fostering the development of those structures, which are recruited for volitional control when vocalizing
2. for prehensile species, vocal control requires greater effort than forelimb/manual control (Koda, Kunieda, & Nishimura, 2018)
3. flyers and swimmers tend to cover much larger areas, and need to communicate across larger distances (Janik and Slater (1997) allude to this)

## 3. Flyers, swimmers, and beyond

This work brings together ecological, anatomical, developmental, and neural data on a range of species (birds, cetaceans, pinnipeds, elephants, bats, rodents, and primates) in an attempt to corroborate the hypothesis proposed. Humans, which are clearly a dexterous *and* vocal learning species, are seemingly an outlier in the relationship I propose here and deserve special attention. A multidimensional approach is followed, according to which there are no magic-bullet explanations for the evolution of particular cognitive abilities. Indeed, this work defends and adds to the notion that the vocal learning capacity and the factors that contribute to it are not all-or-nothing affairs, but rather nuanced (Wirthlin et al., 2019; Martins & Boeckx, 2020; Ravignani & Garcia, 2022).

## Acknowledgements

I thank Cedric Boeckx, Andrea Ravignani, and members of the Cognitive Biology of Language Group (University of Barcelona) for discussion and suggestions on this idea and on other ideas and topics that led to it. This proposal relies in part on work funded by the Portuguese Foundation for Science and Technology (FCT) in the form of a PhD fellowship (grant SFRH/BD/131640/2017).

## References

Fischer, J., Wegdell, F., Trede, F., Dal Pesco, F., & Hammerschmidt, K. (2020). Vocal convergence in a multi-level primate society: Insights into the evolution of vocal learning. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1941), 20202531.

Ghazanfar, A. A., Liao, D. A., & Takahashi, D. Y. (2019). Volition and learning in primate vocal behaviour. *Animal Behaviour*, *151*, 239–247.

Janik, V. M., & Slater, P. J. (1997). Vocal Learning in Mammals. In *Advances in the Study of Behavior* (Vol. 26, pp. 59–99). Elsevier.

Koda, H., Kunieda, T., & Nishimura, T. (2018). From hand to mouth: Monkeys require greater effort in motor preparation for voluntary control of vocalization than for manual actions. *Royal Society Open Science*, *5*(11), 180879.

Martins, P. T., & Boeckx, C. (2020). Vocal learning: Beyond the continuum. *PLOS Biology*, *18*(3), e3000672.

Ravignani, A., & Garcia, M. (2022). A cross-species framework to identify vocal learning abilities in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1841), 20200394.

Wirthlin, M., Chang, E. F., Knörnschild, M., Krubitzer, L. A., Mello, C. V., Miller, C. T., Pfenning, A. R., Vernes, S. C., Tchernichovski, O., & Yartsev, M. M. (2019). A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait. *Neuron*, *104*(1), 87–99.

# EVOLUTION OF TOPICALIZATION IN HUMAN LANGUAGE

DAIKI MATSUMOTO*[1,2]

*matsumoto.daiki.24x@st.kyoto-u.ac.jp
[1]Graduate School of Human and Environmental Studies, Kyoto University
[2]Japan Society for the Promotion of Science

This paper shows that the comparative approach to each subcomponent of our mosaic language capacity can also explain the evolution of seemingly theory-internal specific linguistic phenomena, by taking what is traditionally called "Topicalization" as an example. The first section sets the framework, overviews what Topicalization is with some examples, and introduces the questions to be addressed and answered. The results and hypotheses from the previous analyses of primate calls are sketched in Section 2, which play a significant role in the proposal. Section 3 hypothesizes that the origin of Topicalization can be found in how primate calls are formed. The same section further argues that Self-domestication played a pivotal role in the genesis of Topicalization in human language. Section 4 concludes the paper.

## 1. Introducing the Question to be Asked and Answered

Studies in the evolution of language center around various topics from the hierarchical syntax to the communicative function of our language. As Boeckx (2021) a.o. rightly points out, language is not a monolithic cognitive capacity but a mosaic property of our cognition, consisting of various sub-components. There is mounting evidence in the literature that once this mosaic property is decomposed into pieces, remarkable similarities and relations with other species can be discovered with respect to those subcomponents of our language competence including intentionality (Tomasello 2008), recursive syntax (Hauser & Fitch 2004), semantics (Hurford 2007), and phonology (Samuels 2011).

However, language as a mosaic whole is not entirely limited to those conspicuous cognitive sub-modules. It is widely known that there are many seemingly language-specific semantic, syntactic, phonological and pragmatic phenomena that are considered to be universal across individual languages. One example can be obtained from what is called “Topicalization[1]”. Topicalization is widely known in the relevant literature (see e.g., Culicover, 1999) as a syntactic operation that fronts a “topic” of a sentence to the sentence-initial position:

(1) (Talking about the book over there and your conversation partner asks you about who bought the book when. Then you reply:)
The book over there, John bought (it) when Mary visited him.

In (1), “the book over there” serves as a topic of the conversation as the question by the addressee is about it, and it is fronted to the sentence-initial position in your answer sentence. The rest of the sentence serves as a comment on this topic, which essentially elaborates what this topic is about.

Interestingly, in the same context, (2) sounds infelicitous. This is because “when Mary visited him” is not a topic of the sentence but is a part of the comment about the topic “the book over there” (* indicates that the sentence is ungrammatical).

(2) *When Mary visited him, John bought the book over there.

Furthermore, it is known that (3) is dispreferred, in which the topic is placed at the end of a sentence rather than at the initial position.

(3) *John bought when Mary visited him the book over there.

This talk seeks to show that the nature of Topicalization, which is seemingly language-specific, can also be captured through the lens of the decompositional approach to narrow down the gap between human languages and other species’ cognition. The questions to be asked are: why does Topicalization always “front” a topic rather than postposing etc. the topical element in human language? Is this a purely language-specific phenomenon, the analog of which cannot be observed elsewhere? In what follows, it is argued based upon previous studies (Arnold & Zuberbühler, 2006a, b, 2008, 2012, 2013; Schlenker et al., 2016) that it is not human language specific and the analog can be found in primate calls. We will also see that the fact that Topicalization always fronts a topic finds its evolutionary rationale in how primates form a sequence of alarm calls.

## 2. How Primate Alarm Calls are Pragmatically Formed

---

[1] By “Topicalization,” I mean the grammatical *operation* that yields the result such as the one in (2) in the text.

As is widely known, male Putty-nosed monkeys have two main alarm calls: *pyow*s and *hack*s (see Seyfarth et al., 1980a, b for an initial work in the field). *Pyows* were initially analyzed as calls that signal that a leopard (*Panthera pardus*) is in the vicinity (Arnold & Zuberbühler, 2006a, b). However, in later studies, Arnold and Zuberbühler (2013) modify their previous analysis and propose that *pyow*s should be analyzed as calls that simply draw attention to the presence and location of the caller, signaling the presence of a threat.

*Hack*s were assumed to be indicative of the existence of crowned eagles (*Stephanoaetus coronatus*) nearby (Arnold and Zuberbühler, 2006a, b). However, Arnold & Zuberbühler (2013) has further refined this analysis and proposed the idea that they are true alarm calls that indicate high arousal, which can be triggered not just by an eagle's presence but also by many other phenomena related to aerial threats such as tree fall (Schlenker et al., 2016). These calls facilitate movements appropriate as a reaction to escape the threats.

In what follows, we assume following these authors' later work that *pyow*s are calls that signal the presence of a threat and draw attention of the listeners to the caller while *hack*s signal that there is an alert causing high arousal triggered by the presence of an aerial threat. Below, P stands for *pyow*s and H does for *hack*s.

In their series of work, Arnold & Zuberbühler (2006a, b, 2008, 2012, 2013) have shown that Putty-nosed monkeys sometimes produce distinct $P^+$-$H^+$ sequences. These $P^+$-$H^+$ sequences consist of a small number of Ps followed by a small number of Hs. According to Arnold & Zuberbühler, $P^+$-$H^+$ sequences are analyzed as signals that initiate "group movement".

In sum, the following "semantics" of Putty-nosed monkeys' alarm calls is generalized (from Schlenker et al., 2016 with a slight modification):

(4) The "semantics" of the alarm calls by Putty-nosed monkeys
  a. $P^+$ signals that there is an alert (and it draws attention to the presence and location of the caller).
  b. $H^+$ signals that there is an aerial predator (an eagle and tree fall, to mention just two).
  c. $P^+H^+$ initiates a group movement.

One fact that is particularly noteworthy regarding (4) is that $P^+$ sequences are not observed when an eagle is present in the vicinity. Under the definition in (4a), this absence of $P^+$ is unexpected: since (4a) basically says that $P^+$ can be used if there is an alert and the presence of an eagle is clearly an alert for Putty-nosed monkeys, (4a) expects $P^+$ to be observed in this case, contrary to fact.

To explain this fact while maintaining the crux of the generalization in (4), Schlenker et al. (2016: 15) propose the following principle:[2]

> **(5) Urgency Principle**
> If a sentence *S* is triggered by a threat and contains calls that convey information about its nature and location, no call that conveys such information should be preceded by any call that does not.

Urgency Principle in (5) basically states that calls that convey "more urgent" information should precede ones that convey less urgent one. According to (4), $P^+$ is a call that signals that there is an alert while $H^+$ signals that there is an *aerial* alert. Therefore, (5) predicts that at the presence of an eagle, $P^+$ is preempted by $H^+$. This is what is observed according to the experiments by Arnold & Zuberbühler, and hence Urgency Principle in (5) neatly explains the fact that (4) fails to capture (I refer the reader to Schlenker et al., 2016 and Arnold & Zuberbühler, 2006a, b, 2008, 2012, 2013, for the details of the experiments).

Furthermore, (5) predicts that not only $P^+$ but also $P^+H^+$ is unavailable when there is an aerial alert such as the presence of an eagle or tree fall. According to Schlenker et al. (2016: 17), this is what we observe (again, due to space limitations, I refer the reader to Schlenker et al., 2016 and Arnold & Zuberbühler, 2006a, b, 2008, 2012, 2013, for details).

Summing up this section, the distribution of $P^+$, $H^+$, and $P^+H^+$ is restricted by Urgency Principle defined in (5). Before adopting this idea to the biological underpinning of Topicalization, note in passing that Urgency Principle has an evolutionary rationale: it is plausible to assume that the principle in (5) has emerged evolutionarily since those that obey this principle obviously increase the chance of survival by quickly reacting the potential threats to their lives. Thus, we can claim that it is a residue of group selection.

In the next section, we hypothesize that Urgency Principle is the origin for Topicalization observed in (virtually) every human language.

## 3. Urgency Principle as the Origin of Topicalization

Urgency Principle dictates that those calls that convey information about the nature and location of a threat precede ones that do not convey such information.

---

[2] Of note here is that by this principle Schlenker et al. seek to propose a compositional semantic account of those alarm calls. Since their compositional semantics does not play a role in what follows and this principle still plays its role even if their compositional semantics is abstracted away from, this paper does not delve into the details of Schlenker et al.'s semantic account.

In the wild world of the animal kingdom, this abstractly means that calls that convey information "about their lives" precede ones that do not convey such information. Given the fact that those calls further evoke escaping movement of those addressees, this can be further interpreted as follows: those calls uttered in accord with (5) "are about" the comment (i.e., types of movement that they evoke).

Under this interpretation of the pragmatic function of Urgency Principle, the similarity between the principle and human language Topicalization becomes obvious. Both "front" the information that conveys what the comment is about.

Based upon this, this article hypothesizes the following:

(6) Topicalization is evolutionarily derived from Urgency Principle.

(6) claims (i) that Topicalization, which is seemingly not just human-specific but also language-specific, can find its analog in how primate calls are shaped in a particular manner, and (ii) that it has Urgency Principle as its evolutionary basis. Below, the rationales for this hypothesis are provided.

Of note first is that Topicalization has a communicative function and it plays little role (if any) in constructing a propositional meaning of a sentence: even if Topicalization does not happen, the basic semantics of (2) remains the same, and its primary role is to front a topic of the current discourse and create a structure in which the topic and the comment about it are divided in a communicatively efficient fashion. This suggests that Topicalization is not required for semantic composition and hence should not be observed in the first place if no communicative function exists in our language use. Succinctly put, Topicalization primarily has a communicative function.

Notice at this point that Urgency Principle is also a principle that forms a sequence that conveys important/urgent information in an efficient way in accord with the external condition under which those calls are produced.

From these, it is not unreasonable to assume that the functional basis of Topicalization can be found in Urgency Principle. Then, how did Topicalization evolve from Urgency Principle? Notice already that Topicalization in human language is not triggered by urgency in any sense.

To address this issue, I hypothesize that human self-domestication played a key role in the evolution of Topicalization from Urgency Principle. As is widely discussed in the relevant literature (see Theofanopoulou et al., 2017 for an excellent overview and insights), modern humans are claimed to have undergone a process of self-domestication which caused the globularization of the brain case and a reduction in tooth size among many other features peculiar to domesticated species (see also Wilkins, Wrangham, & Fitch, 2014 for some pioneering work in the field). It has also been widely argued that self-domestication reduced reactive aggression in humans. Self-domestication in humans is argued to be triggered by many types of socio-environmental chaos such as the climate catastrophe during the Last Glaciation, which facilitated prosocial behavior of our ancestors (Spikins et al., 2021). And this is claimed to have resulted in a selection for less emotionally reactive group partners and for receptiveness towards extra-group individuals (Hare, Wobber, & Wrangham, 2012; Pisor & Surbeck, 2019).

Recently, Benítez-Burraco, Ferretti, & Progovac (2021: 1) argue that "the reduction in reactive aggression, one of the key factors in self-domestication processes, enabled us to fully exploit our cognitive and interactional potential as applied to linguistic exchanges," claiming that it played a decisive role in the evolution of our pragmatic capacities.

Based upon this illuminating previous work, I hypothesize that Topicalization, which has a clear communicative (and hence, pragmatic) function, became available to humans through self-domestication via socialization, emancipating us from the prison of urgent threats and reactive aggression.[3] In other words, because of self-domestication, Urgency Principle could be safely applied to our language use with no urgency. Once it became nonurgent, it has a clear communicative function that forms a signal sequence that conveys information in an efficient way in accordance with the discourse, which eventually culturally evolved and stabilized in the species along with other pragmatic capacities.

As one of the anonymous reviewers suggests, to which I am indebted, Urgency Principle in (5) can be further restated as a principle that requires "calls that contain more information that is currently relevant come first." Under this abstracted interpretation of (5), the similarity between (5) and Topicalization should be fairly obvious. Based upon this suggestion by the reviewer, I submit that what self-domestication facilitated was to liberate our ancestors from urgency, and it allowed us to generally front more relevant information in the discourse.

In sum, the hypothesis is that though Topicalization resulted from Urgency Principle, it became nonurgent due to self-domestication. An immediate virtue of this hypothesis is that it can be further tested experimentally, by designing a

---

[3] One of the reviewers wonders whether the current proposal assumes that "Topicalization emerged fairly late in human language evolution and so that self-domestication episode is the most relevant one". I indeed hypothesize so. This can be tested experimentally via simulation by letting a language evolve and seeing in what stage(s) Topicalization (or a phenomenon analogous to it) emerges. I leave this important work for future research.

computer simulation which examines whether liberation from urgency indeed facilitates the generalization of (5) as a condition of relevance.

## 4. Conclusion

To the best of my knowledge, specific grammatical operations have thus far escaped serious investigation in the circle of language evolution. This paper has brought up the operation called "Topicalization" to the body of research questions in the field of evolution of language and has sought to show that the comparative approach that has been proven to be fruitful in this field sheds light on how this specific, seemingly language-specific operation arose in our lineage.

If language evolution is about how our language biologically evolved, then those specific linguistic operations/phenomena should also be taken as serious objects of evolutionary inquiry. This paper is an attempt to show with a specific example that those theory-internal linguistic operations that have hitherto escaped serious evolutionary attention can be studied in a basic "divide and conquer," comparative fashion. Needless to say, there are scores of other specific grammatical operations/phenomena discussed neither in this paper nor in the previous evolinguistic literature. It is hoped that those will also be studied in the comparative way, along with search for neurological underpinning among others, in future research.

### *Acknowledgements*

I thank Koji Fujita for a series of discussions on this issue. This work is supported by JSPS KAKENHI Grant Number JP20J20039.

## References

Arnold, K., & Zuberbühler, K. (2006a). The alarm calling system of adult male putty-nosed monkey *Cercopithecus nictitans martini*. *Animal. Behavior*, 72, 643-653.

Arnold, K., & Zuberbühler, K. (2006b). Semantic combinations in primate calls. *Nature* 441, 303.

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology* 18(5), R202-R203.

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain Language.* 120(3), 303-309.

Arnold, K., & Zuberbühler, K. (2013). Female putty-nosed monkeys use experimentally altered contextual information to disambiguate the cause of male alarm calls. *PLoS One* 8(6), e65660. doi: http://dx.doi.org/10.1371/journal.pone.0065660.

Benítez-Burraco, A., Ferretti, F., & Progovac, L. (2021). Human self-domestication and the evolution of pragmatics. *Cognitive Science* 45, e12987. doi: 10.1111/cogs.12987.

Boeckx, C. (2021). *Reflections on language evolution: From minimalism to pluralism*. Berlin: Language Science Press.

Culicover, P. W. (1999) *Syntactic nuts*. Oxford: Oxford University Press.

Elordieta G., Lee, C., Gordon M., & Büring D. (eds.) (2007). *Topic and focus: Cross-linguistic perspectives on meaning and intonation*. Netherlands: Springer.

Hare, B., Wobber, V., & Wrangham, R. (2012). The self-domestication hypothesis: Evolution of bonobo psychology is due to selection against aggression. *Animal Behavior*, 83(3), 573-585.

Hauser, M. D., & Fitch, W. T. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656), 377-380.

Hurford, J. (2007). *The origins of meaning*. Oxford: Oxford University Press.

Pisor, A. C., & Surbeck, M. (2019). The evolution of intergroup tolerance in nonhuman primates and humans. *Evolutionary Anthropology*, 28(4), 210-223.

Samuels, B. D. (2011). *Phonological structure: A biolinguistics approach*. Oxford: Oxford University Press.

Schlenker, P., Chelma, E., Arnold, K., & Zuberbühler, K. (2016). *Pyow-hack* revisited: Two analyses of Putty-nosed monkey alarm calls. *Lingua*, 171, 1-23.

Seyfarth, R.M., Cheney, D. L., & Marler, P. (1980a). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, 210, 801-803.

Seyfarth, R.M., Cheney, D. L., & Marler, P. (1980b). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Animal Behavior*, 28, 1070-1094.

Spikins, P., French, J. C., John-Wood, S., & Dytham, C. (2021). Theoretical and methodological approaches to ecological changes, social behaviour and human intergroup tolerance 300,000 to 30,000 BP. *Journal of Archaeological Method and Theory*, 28(1), 53-75. doi: https://doi.org/10.1007/s10816-020-09503-5.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Theofanopoulou, C., Gastaldon, S., O'Rourke, T., Samuels, B. D., Messner, A., Martins, P. T., et al. (2017). Self-domestication in Homo sapiens: insights from comparative genomics. *PLoS One* 12: e0185306. doi: http://dx.doi.org /10.1371/journal.pone.0185306.

Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The "domestication syndrome" in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics*, 197(3), 795-808. doi: https://doi.org/10.1534/genetics.114.165423.

# CO-EVOLUTIONARY LINKS BETWEEN LINGUISTIC ALIGNMENT AND COOPERATION – AN EXPERIMENTAL STUDY

THERESA MATZINGER*[1,2], MAREK PLACIŃSKI[1], ADAM GUTOWSKI[1],
MARIUSZ LEWANDOWSKI[1], PRZEMYSLAW ZYWICZYNSKI[1] and SŁAWOMIR WACEWICZ[1]

*Corresponding Author: theresa.matzinger@univie.ac.at
[1]Centre of Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland
[2]Department of English, University of Vienna, Vienna, Austria

Linguistic communication requires a high degree of cooperation between interaction partners (Tomasello, 2008; Fitch, 2010; Grice, 1975; Hurford, 2007; Knight, 2016). At the same time, language unlocks mechanisms for maintaining and strengthening cooperation: it facilitates cooperation through verbal coordination (Gärdenfors, 2004), or raises the costs of non-cooperation by spreading the track record of individuals via gossip (Dunbar, 1996). These close links between language and cooperation make it highly plausible that these two traits co-evolved during their evolutionary emergence. Our study explores these co-evolutionary links between cooperation and language and helps to shed light on possible causalities and directionalities of these links.

In particular, we hypothesize that people are more likely to cooperate with individuals that align with them linguistically, i.e. share lexical or syntactic choices (Pickering & Garrod, 2004). Such correlations have already been observed with respect to task success (Reitter & Moore, 2014), and we expect that the same positive relationship exists between language and cooperation.

To test this, we conducted an online experiment with 40 native English-speaking participants. Each participant communicated with two interaction partners in a picture-naming task (Bock, 1986), in which the participants and their partners took turns describing pictures that represented ditransitive events. After an initial communication stage, where both partners (which were in fact standardized chat bots) did not align their syntactic constructions with the participants, one of the partners switched to an aligned syntactic construction, whereas the other partner continued to use a non-aligned syntactic construction (Fig. 1). After this communicative interaction, participants had to decide in a two-alternative forced choice task with which of the two partners they would like to play a cooperation game that determined the financial bonus that they would receive for participation. The participants' choices in this task were our main variable of interest. We predicted that linguistically aligning communication partners would be preferred as cooperation partners over non-aligning ones.

Our results did not confirm this prediction. Instead, aligning (47.5%) and non-aligning (52.5%) partners were chosen equally often as cooperation partners (confidence intervals include 50%). Further exploratory analyses showed that there was no correlation between the attention that participants paid to the language used by their partners and their partner choices ($\chi^2 = 0.17$, $p = 0.68$, $\varphi = 0.07$): also when participants reported that they actively paid attention to and based their decisions on the "correctness" of their partners' responses, they did not choose the aligned partners significantly more often than the unaligned ones (note that "correctness" is the participants' subjective impression, since both aligned and non-aligned partners answered grammatically correctly).

Our study did not find evidence that syntactic alignment is a decisive factor for people when choosing their cooperation partners. Other yet unexplored factors may override the influence of linguistic alignment in such decisions. For example, people may avoid partners who mimic them exactly, and rather regard those using different syntactic structures as more competent, honest or trustworthy partners in cooperative interactions. However, since our design relied on only a single instance and a single type of alignment (syntactic alignment), it might not have been sensitive enough to capture subtle effects. Follow-up research is necessary, in particular with alignment unfolding over a longer course of interaction, to reach a firmer conclusion regarding the interdependence, and possible co-evolution, of linguistic alignment and cooperation.

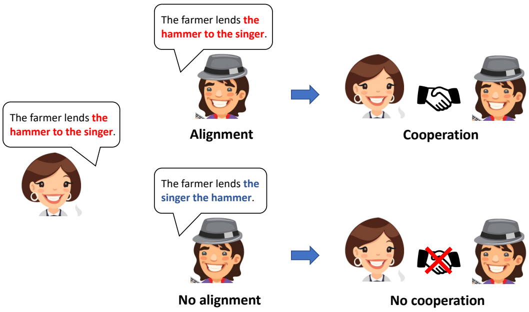


Figure 1. Experimental design and hypotheses.

## References

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.

Dunbar, R. I. M. (1996). *Grooming, gossip and the evolution of language*. London: Faber and Faber.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Gärdenfors, P. (2004). Cooperation and the evolution of symbolic communication. In K. Oller & U. Griebel (Eds.), *The Evolution of Communication Systems* (pp. 237–256). Cambridge, MA: MIT press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). New York: Academic Press.

Hurford, J. R. (2007). *The origins of meaning: Language in the light of evolution I* (Vol. 8). Oxford and New York: Oxford University Press.

Knight, C. (2016). Puzzles and mysteries in the origins of language. *Language & Communication*, 50, 12–21.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27(2), 169–226.

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, 76, 29–46.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

# AESTHETIC PERCEPTION OF PROSODIC PATTERNS AS A POTENTIAL FACTOR IN THE CULTURAL EVOLUTION OF LANGUAGE

THERESA MATZINGER*[1,2,3], EVA SPECKER[4], NIKOLAUS RITT[1] and TECUMSEH FITCH[2]

*Corresponding Author: theresa.matzinger@univie.ac.at
[1]Department of English, University of Vienna, Vienna, Austria
[2]Department of Behavioral and Cognitive Biology, University of Vienna, Vienna, Austria
[3]Centre of Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland
[4]Department of Basic Psychological Research, University of Vienna, Vienna, Austria

All human cultures appreciate art and can perceive visual, verbal, or musical stimuli in terms of their aesthetic appeal. For example, aesthetic preferences regarding prosodic patterns play a prominent role for the appeal of poetry (Nadal & Vartanian, 2019; Rastall, 2008). If such preferences also apply in the perception of spontaneous everyday speech, they may pose a constraint on language change (Rastall, 2008): we hypothesize that aesthetically appealing linguistic features are learned easily and used frequently, and will thus be culturally transmitted to future generations of speakers more successfully than less appealing features (cf. Smith & Kirby, 2008).

In our exploratory study, we investigated a crucial baseline for this hypothesis, namely if there were indeed differences in listener's aesthetic judgements of linguistic features. Specifically, we focused on the aesthetic perception of temporal rhythmic patterns in polysyllabic words. On the one hand, words might be regarded as most aesthetic if their syllables are isochronous because isochrony has a facilitatory effect on auditory processing, and people have a general propensity for regular patterns (Ravignani & Madison, 2017). On the other hand, listeners may also perceive irregular patterns as aesthetically appealing (Westphal-Fitch & Fitch, 2013). In that case, words with deviations from isochrony might be judged as more pleasing than isochronous stimuli.

To explore the potential link between words' rhythmic patterns and aesthetic perception, we tested 180 native-German-speaking participants on their aesthetic evaluation of artificially generated trisyllabic pseudo-words. Each participant made valence ratings of 20 words that were each presented in 3 different conditions in a random order: a) with isochronous syllables, b) with the initial, medial or final syllable lengthened and c) with the initial, medial or final syllable shortened by 50% of its original duration (400 ms). Each participant ranked each word three times, namely on its 'likability', on its 'beauty' and on its 'naturalness', which together served as indicators of 'aesthetic appeal'.

Cumulative Link Mixed Models revealed that, overall, isochronous syllables were preferred over deviations of isochrony. Especially, shortened syllables had a prominent negative effect on aesthetic appeal. The only modification that participants judged as slightly more aesthetically appealing than isochrony was word-final lengthening. These results were similar for 'likability', 'beauty' and 'naturalness'.

The positive ratings of rhythmic patterns are unlikely to have been influenced by their occurrence frequencies in the participants' native language (Bybee, 2007) because word-medial syllables, which are typically stressed and thus lengthened in German (Domahs, Plag, & Carroll, 2014) have not been evaluated as aesthetically appealing when lengthened in our experiment.

Interestingly, the aesthetic appeal of prosodic patterns in our study corresponded to their effectiveness for speech segmentation in other experiments, where words with finally lengthened syllables could be extracted from continuous speech more successfully than words with finally shortened syllables (Matzinger, Ritt, & Fitch, 2021). Together, these findings indicate a potential connection between aesthetics and language learning. Thus, overall, this study serves as an important starting point for testing the role of aesthetic perception of linguistic input for the cultural evolution of linguistic patterns.

Further research should test the role of aesthetic appeal in language change more directly, for example in iterated learning experiments (Kirby, Cornish, & Smith, 2008). Also, future research should consider that aesthetic preferences are likely to arise from a combination of factors, including the occurrence frequencies of the target patterns, their position withing a sentence, or an interplay of familiarity and novelty biases (Sluckin, Hargreaves, & Colman, 1983).

## References

Bybee, J. (2007). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301571.001.0001

Domahs, U., Plag, I., & Carroll, R. (2014). Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics*, *17*(1), 59–96. https://doi.org/10.1007/s10828-014-9063-9

Kirby, S., Cornish, H., Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Matzinger, T., Ritt, N., & Fitch, W. T. (2021). The Influence of Different Prosodic Cues on Word Segmentation. *Frontiers in Psychology*, *12*(March). https://doi.org/10.3389/fpsyg.2021.622042

Nadal, M., & Vartanian, O. (Eds.). (2019). *The Oxford Handbook of Empirical Aesthetics*. Oxford: Oxford University Press.

Rastall, P. (2008). Aesthetic responses and the “cloudiness” of language: Is there an aesthetic function of language? *La Linguistique*, *44*, 103–132. https://doi.org/10.3917/ling.441.0103

Ravignani, A., & Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology*, *8*, 1–13. https://doi.org/10.3389/fpsyg.2017.01820

Sluckin, W., Hargreaves, D. J., & Colman, A. M. (1983). Novelty and human aesthetic preferences. In J. Archer & L. I. A. Birke (Eds.), *Exploration in animals and humans*. Wokingham: Van Nostrand Reinhold, 245–269.

Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1509), 3591–3603. https://doi.org/10.1098/rstb.2008.0145

Westphal-Fitch, G., & Fitch, W. T. (2013). Spatial Analysis of “Crazy Quilts”, a Class of Potentially Random Aesthetic Artefacts. *PLoS ONE*, *8*(9). https://doi.org/10.1371/journal.pone.0074055

# DISENTANGLING THE STATISTICAL PROPERTIES THAT DRIVE LANGUAGE ACQUISITION: EVIDENCE FROM MAXIMALLY DIVERSE LANGUAGES

Jekaterina Mažara*[1,2], Giuachin Kreiliger[1,2], Sabine Stoll[1,2], and Balthasar Bickel[1,2]

*Corresponding Author: jekaterina.mazara@uzh.ch
[1]Department of Comparative Language Science University of Zurich, Switzerland
[2]Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

A key property of the human language faculty is that any of its instantiations can be acquired by any child in roughly the same time frame. A powerful suggestion for what makes this possible is the idea that children can rely on statistical properties for segmenting and learning elements such as morphemes or words (Saffran, 2003; Saffran, Aslin, & Newport, 1996; Gerken, 2005; Pelucchi, Hay, & Saffran, 2009; Hay, Pelucchi, Estes, & Saffran, 2011; Thiessen, 2011; Höhle, 2015). However, an unresolved question is which specific aspects of statistical distributions drive or differentiate the acquisition process in real-world data: is it the sheer number of items that need to be learned? Or distributional differences across structural or semantic types, or across entire languages? Or frequency classes?

Here we address these questions by assessing potential factors that might influence the acquisition of verb forms in morphologically highly diverse languages. Our data consists of longitudinal recordings of 20 children and their surrounding speakers from five naturalistic corpora from the *ACQDIV* data base (Moran, Schikowski, Pajović, Hysi, & Stoll, 2016): Chintang (Sino-Tibetan), English (Indo-European), Japanese (Japonic), Turkish (Turkic), and Yucatec (Mayan). As potential factors we tested the following: (i) **the difference between *heads* and *dependents***. *Heads* include the main stem of a verb form that carries structurally independent information, while *dependents* comprise elements that provide modifying lexical, derivational, and grammatical information, such as affixes and secondary stems (Stoll, Mazara, & Bickel, 2017). We hypothesize that heads are easier to learn than dependents because they are cognitively more salient. (ii) **the difference between languages**, hypothesizing that some languages are generally acquired faster than others. (iii) the **deviation of the empirical input distributions from the theoretical Zipf distribution** (quantified by the contribution of a parabolic function to fitting the log-log rank-frequencies), hypothesizing that convexity in rank-frequency distributions facilitates learning because there are more high-frequency items than expected. (iv) the **log number of distinct heads and**

**dependents**, assuming that a higher number would indicate a more difficult system for learners. (v) the **entropy** of items (measured with the Chao-Shen estimator (Chao & Shen, 2003) to account for possible unseen elements), hypothesizing that lower entropies facilitate learning (Lavi-Rotbain & Arnon, 2022).

To evaluate the children's development of productivity in verb form use over time, we computed the (log) ratio in usage entropies between the target child and surrounding adult speakers within each recording session (Stoll et al., 2012).

We applied a hierarchical nonlinear scale/location ("distributional") model (Bürkner, 2018) to fit the development of the log entropy ratios over time using the exponential function to describe the learning curves, which allows us to evaluate the speed of acquisition. To allow for variation at an individual level, we model the children as random effects. We fitted models with one of the five factors each (plus an ancillary model of item type and language together) and compared them through model stacking, estimating the relative weight of each factor in predictive performance during leave-one-out cross-validation (Vehtari, Gelman, & Gabry, 2017).

Model stacking (Yao, Vehtari, Simpson, & Gelman, 2018) shows that the data are best predicted by the convex deviation from the theoretical Zipfian distribution, i.e. distributions with more higher-frequency types faciliate learning and increase the rate of acquisition. The (log) number of items has a slightly weaker but still appreciable impact (stacked prediction model weight .45 vs .5 for the deviation from the Zipfian distribution). All other factors predict the data much less well (receiving no weight in stacking), i.e., differences between languages, item types, and entropies have no impact on the acquisition process.

Our findings suggest that the children in our sample indeed do make use of statistical properties of the input during learning, and that it is specifically the number of high-frequency items and the total number of items that matters.

## References

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*, 395–411.

Chao, A., & Shen, T.-J. (2003). Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics*, *10*(4), 429–443.

Gerken, L. (2005). Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67–B74.

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93–106.

Höhle, B. (2015). Crosslinguistic perspectives on segmentation and categorization in early language acquisition. In E. L. Bavin & L. R. Naigles (Eds.), *The cambridge handbook of child language* (p. 159-182). Cambridge University

Press.

Lavi-Rotbain, O., & Arnon, I. (2022). The learnability consequences of zipfian distributions in language. *Cognition*, *223*, 105038.

Moran, S., Schikowski, R., Pajović, D., Hysi, C., & Stoll, S. (2016). *The acqdiv corpus.*

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926 - 1928.

Stoll, S., Bickel, B., Lieven, E., Banjade, G., Bhatta, T. N., Gaenszle, M., Paudyal, N. P., Pettigrew, J., Rai, I. P., Rai, M., & Rai, N. K. (2012). Nouns and verbs in chintang: children's usage and surrounding adult speech. *Journal of Child Language*, *39*, 284-321.

Stoll, S., Mazara, J., & Bickel, B. (2017). The acquisition of polysythetic verb forms in chintang. In M. Fortescue, M. Mithun, & N. Evans (Eds.), *Handbook of polysynthesis* (p. 495-517). Oxford University Press.

Thiessen, E. D. (2011). Domain-general constraints on statistical learning. *Child Development*, *82*(2), 462–470.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, *27*(5), 1413–1432.

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average bayesian predictive distributions (with discussion). *Bayesian Analysis*, *13*(3), 917–1007.

# THE EMERGENCE OF GAZE CONVENTIONS IN INTERPERSONAL COORDINATION

Gregory Mills*, Remko Boschker

*Corresponding Author: g.j.mills@rug.nl
Centre for Language and Cognition Groningen,
University of Groningen, Netherlands

A central finding in psycholinguistic research is that when people interact with each other they rapidly develop new idiosyncratic referring expressions and associated meanings with their conversational partner (Clark, 1996; Pickering and Garrod, 2004)

In addition to natural language, conversational interaction is underpinned by myriad non-verbal signals which are used, inter-alia, to regulate turn-taking (Argyle, 1988; Holler et al., 2016). For example, speakers tend to look away from their addressee while speaking and then re-establish eye-contact to signal the end of their turn.

However, it is currently unclear whether such turn-taking signals are static and fixed, or whether, like natural language, they are negotiated by participants during interaction. To address this question participants play a novel collaborative task, in virtual reality. The task is played by 3 participants, and is inspired by games such as guitar hero and dance-dance revolution. The three key differences are: (1) Instead of performing target sequences of musical notes or dance moves, participants need to perform sequences of gaze events. E.g. a typical target sequence might be: "*Person 2 must look at person 3. Then person 3 must look at person 1. Then person 1 and person 2 look at each other*" (2) On each trial, only one participant (the director) sees the target sequence. This means that in order for the group to complete the target sequence, the director has to instruct the others (3) Crucially the participants are not allowed to use natural language – they may only communicate by looking at each other. Solving the task, therefore, requires that participants bootstrap a communication system, solely using their gaze patterns.

We conducted a set of experiments where triads played this game for 25 minutes followed by a manipulation which swapped participants` identities: In Participant 1`s view, Participant 2`s avatar is controlled by Participant 3, while Participant 3`s avatar is controlled by Participant 2. In Participant 2`s view, Participant 1`s avatar is controlled by Participant 3 and vice versa. Similarly for Participant 3`s view. The intention behind this manipulation is that if participants establish pair-specific routines with each of their partners, this swap will disrupt co-ordination, since after the swap, participants will be interacting with a different partner who is using different routines.

The results show that during the first 25 minutes of the task, triads are able to establish a communication system that allows them to solve sequences of up to 8 gaze events. Moreover, the results show that after swapping participants` identities, triads perform worse at the task, suggesting that the routines established to coordinate the timing and sequencing of eye-gaze can be flexibly negotiated during interaction.

## References

Argyle, M. (1988). Bodily Communication. 2nd Edition. Methuen, London

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Holler, J., Kendrick, K. H., Casillas, M., & Levinson, S. C. (2016). Turn-taking in human communicative interaction. Frontiers Media SA.Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707-784.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. Behavioral and brain sciences, 27(2), 169-190.

# SELF-REPAIR INCREASES ABSTRACTION OF REFERRING EXPRESSIONS

Gregory Mills*, Gisela Redeker

*Corresponding Author: g.j.mills@rug.nl
Centre For Language and Cognition, Groningen, University of Groningen

When interlocutors repeatedly describe referents to each other, they rapidly converge on referring expressions which become increasingly systematized and abstract as the interaction progresses. This occurs for a wide range of referents, e.g. when referring to spatial locations (Garrod and Doherty, 1994; Roberts et al., 2016; Mills, 2014), music (Healey et al., 2007), concepts (Schwartz, 1995), confidence (Fusaroli et al., 2012), and temporal sequences (Mills, 2011). Cumulatively, these findings suggest that interaction in dialogue places important constraints on the semantics of referring expressions.

However, there is currently no consensus about how best to account for how convergence develops. The iterated learning model (Kirby, Griffiths and Smith, 2014) explains convergence as arising out of individual cognitive biases; the interactive alignment model of Pickering and Garrod (2004; 2021) favours alignment processes, while the collaborative model of Clark (1996) emphasizes the role of positive feedback. By contrast, Healey et al., (2007) argues that negative evidence of understanding plays the central role: When interlocutors initiate repair, this allows them to interactively identify, diagnose and resolve any differences in interpretation between them and their conversational partner. Addressing these differences accelerates convergence.

To investigate in closer detail how negative evidence contributes toward convergence, we report a variant of the “maze task” (Pickering and Garrod, 2004). Participants communicate with each other via an experimental chat tool (Healey and Mills, 2006), which automatically transforms participants' private turn-revisions into public self-repairs that are made visible to the other participant. For example, if a participant, A types:

**A:** Now go to the square on the left, next to the big block on top

and then before sending, A revises the turn to:

**A:** Now go to the square on the left, next to the third column

The chat server automatically detects the revised text and inserts a hesitation marker (e.g. "umm" or "uhhh" immediately preceding the revision). This would yield the following turn, sent to B:

**A:** Now go to the square on the left next, to the big block on top umm.. I meant next to the third column

Participants who received these transformed turns used more abstract and systematized referring expressions, and also used a larger vocabulary (i.e. more unique words). However, dyads who received the interventions solved fewer mazes and produced more turns than participants in the control group.

We argue that this effect is due to the artificial self-repairs causing participants to put more effort into diagnosing and resolving the referential coordination problems in the task. At the start of the experiment this leads to fewer solved mazes, but ultimately yields better grounded spatial semantics and consequently leads to increased use of abstract referring expressions.

## References

Clark, H. H. (1996). *Using language*. 1996. Cambridge University Press: Cambridge

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms quantifying the benefits of linguistic coordination. *Psychological Science*, 0956797612436816.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181-215.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292-304.

Healey, P. G., Swoboda, N., Umata, I., & King, J. (2007). Graphical language games: Interactional constraints on representational form. *Cognitive Science*, 31(2), 285-30

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28, 108-114 https://doi.org/10.1016/j.conb.2014.07.014.

Mills, G. J. (2011). The emergence of procedural conventions in dialogue. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Mills, G. J. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology*, 32, 158-173.

Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169-190.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52-66.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences*, 4(3), 321-354.
.

# EFFECT OF ENVIRONMENTAL RICHNESS ON RESPONSIVENESS OF RATS TO SPECIES-SPECIFIC EMOTIONAL CALLS

Marika Minamii*[1,2], Yumi Saito[3], and Kazuo Okanoya[2,1]

*Corresponding Author: mnmmrk-1066@g.ecc.u-tokyo.ac.jp

[1] Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

[2] Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan

[3] Animal Science and Biotechnology, Azabu University, Kanagawa, Japan

Language and speech communication are closely intertwined, and a comprehensive understanding of the evolution and origin of language cannot be achieved without the study of sound communication in a variety of animal species. Rats produce two types of ultrasonic vocalization (USV) in response to emotional and social contexts. Previous studies have suggested that 50 kHz "pleasant" USV increases heart rate with a significant tendency, while 22 kHz "unpleasant" USV decreases heart rate (Saito, 2019). However, the environment in which laboratory rats are raised is quite different from the environment in which wild rats live. Brydges *et al.* (2011) suggested that rats raised in rich environments tend to judge intermediate stimuli as positive stimuli. It has also been shown that environmental enrichment stimulates progenitor cell proliferation in the amygdala (Okuda et al., 2009). Based on these studies, because the laboratory rearing environment may alter the innate response to USVs, it is necessary to conduct further studies with more wild-like enrichment environments. Therefore, in this study, we made a hypothesis that rats reared in a rich environment respond less to 22 kHz USV or more to 50 kHz USV and examined whether the environmental richness affects the responsiveness to USV.

In this study, rats raised in an enriched environment were exposed to USV, and their responses to USV were examined in comparison with those of rats raised in a standard environment. Following Saito's study (2019), the responses to USV were indexed by increases or decreases in heart rate. If the rats raised in the enriched environment are more optimistic and the decrease in heart rate when listening to 22 kHz USV is not significant, or if the heart rate of rats listening to 50 kHz USV is significantly increased, it is assumed that the enrichment altered the rats' responsiveness to USV.

Of the 12 weaned male rats, 6 were pair-reared in cages (approximately 25 × 40 × 20 cm) lined with recycled paper bedding. The others were kept in large cages (approximately 40 × 65 × 30 cm) with a thick layer of pepa-kurin (JapanSLC, *Shizuoka*, *Japan*) and several enrichment supplies, three animals per cage. Two months later, 50 kHz and 22 kHz USV were presented, and heartbeats were measured during listening.

MATLAB was used for the following analysis. The baseline was defined as the time up to 7 seconds before the start of voice presentation, and the test was defined as 7 seconds after. The RRI, the interval between R waves, was calculated, and the mean of the RRI (mean RRI) was calculated as the heart rate interval. Prolongation of the RRI indicates a falling heart rate and shortening of the RRI indicates a rising heart rate (Jose, 1966). For these indices, baseline values were subtracted from the tests in the block. For example, if the difference is significantly greater than zero, we can say that listening to the USV decreased heart rate.

The following analysis was then conducted using R. To examine whether each index changed from baseline due to the presentation of USV, we tested the median values by one-sample Wilcoxon signed rank test and found no significant changes in any of the conditions. The test results were shown in Fig. 1.

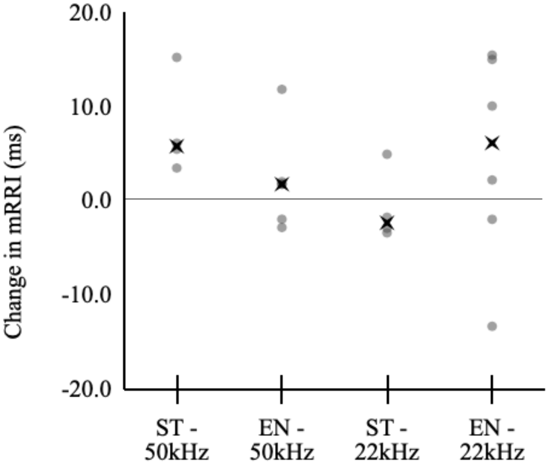


Figure 1. Change in mRRI from baseline. ST: standard group (N = 4), EN: enriched group (N = 6), ×: the median.

In conclusion, both the environmental richness and the category of USV did not affect the response of the autonomic nervous system in rats. However, some individuals did not return their heart rate to baseline after listening to the USV, which may have prolonged the effect of the preceding stimulus. Thus, we need to review whether the baseline and interstimulus interval settings were appropriate. Moreover, we would like to investigate whether the enriched environment operation was appropriate through further experiments.

## References

Brydges, N. M., Leach, M., Nicol, K., Wright, R., & Bateson, M. (2011). Environmental enrichment induces optimistic cognitive bias in rats. *Animal Behaviour*, 81(1), 169-175.

Okuda, H., Tatsumi, K., Makinodan, M., Yamauchi, T., Kishimoto, T., & Wanaka, A. (2009). Environmental enrichment stimulates progenitor cell proliferation in the amygdala. *Journal of neuroscience research*, 87(16), 3546-3553.

Saito, Y. (2019). Categorical perception and processing of emotional calls: The study of rat ultrasonic vocalizations. [Doctoral dissertation, The University of Tokyo]. (In Japanese.)

# LINGUISTIC LAWS IN PRIMATE VOCAL COMMUNICATION

Steven Moran*[1,2], Quentin Gallot[1], Auriane Le Floch[1,3], Marco Maiolini[1], Adrian Soldati[1,4], and Klaus Zuberbühler[1,4]

*Corresponding Author: steven.moran@unine.ch
[1]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[2]Department of Anthropology, University of Miami, Coral Gables, USA
[3]Institut des Sciences Cognitives Marc Jeannerod, CNRS, Lyon, France
[4]School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK

Researchers have increasingly focused on the commonalities between animal communication and language with the aim to identify what makes language unique and to shed light on how it evolved. One area of recent study is whether linguistic laws also characterize vocal communication in other animals. Given their relevance for the evolution of language, here we provide a comprehensive review of the presence or absence of linguistic laws reported in studies on primate vocal communication. We find that primate vocal production follows Menzerath's law, while adherence to the other laws is mixed. Moreover, we raise three important points to consider when studying linguistic laws: the role of sexual dimorphism in vocal production, the criteria used to define a vocal sequence, and the choice of vocal units for analysis. Thus, this review provides a road map for future studies investigating linguistic laws in primates and is aimed at making results more comparable across species and signals.

## 1. Introduction

In the last two centuries linguists and biologists have increasingly focused on the commonalities between animal communication systems and language. The aim is to identify what makes language unique and to shed light on its origin.

The best known account is Hockett's (1960) universal design features of human language. Hockett used them to distinguish between animal communication systems and language, noting that most features (originally 9 of 13) are attested in other species. Ever since Hockett proposed the design features, they have been used to juxtapose human language with communication systems of other animals for descriptive purposes, even though the study of language evolution itself has evolved in the last sixty years. Their use is limited as an evolutionary theoretical framework (Wacewicz & Żywiczyński, 2015). Furthermore, the increased access to large datasets and machine learning has given rise to the application of quantitative and information-theoretic approaches in biological systems more generally.

We provide a comprehensive review of the current state of the art of linguistic laws reported in primate vocal communication. We note in which species, signals, and contexts that there is support for the presence or absence of different linguistic laws. Our work builds on that of Semple, Cancho, and Gustison (2021) and we provide an online and openly accessible table of the current reported findings. These results shed light on, and provide a road map for, future studies on primates and other animal species to investigate communication systems, linguistic laws, and language evolution.

## 2. Linguistic laws

Linguistic laws are mathematically formulated statistical regularities discovered by linguists studying properties of (written) language. These patterns appear at various linguistic levels (e.g., sounds in words, words in text) and their existence is often attributed to universal principles of efficiency in human language (cf. Levshina and Moran (2021)). There are four well-studied linguistic laws.

The most well-known is referred to as Zipf's law or Zipf's principle of least effort (Zipf, 1936). It states that the frequency distribution of words in a language is inversely proportional to their frequency rank. Zipf theorized that this distribution of words is due to the principle of least effort, i.e., speakers minimize their effort when transmitting messages (hence why the most frequent words in a language tend to also be the shortest). Work by Yu, Xu, and Liu (2018) shows that Zipf's law holds across a sample of 50 languages. However, why the Zipfian distribution – one of several power law distributions – occurs is still not well understood. Power law distributions are found in numerous biological, physical, and human-made phenomena, and they emerge naturally from latent variables (Aitchison, Corradi, & Latham, 2016) or from almost nothing (Piantadosi, 2014).

Zipf's law of abbreviation (Zipf, 1949), also known as the law of brevity, states that more frequently used words will tend to be shorter than less frequently used words within a language. This law in particular captures an aspect of the efficiency of language use (Kanwal, Smith, Culbertson, & Kirby, 2017).

Menzerath-Altmann's law states that the size of constituents within a linguistic construction decrease in size as constituents get longer, and vice versa (Menzerath, 1954; Altmann, 1980). For example, longer words tend to have shorter syllables and shorter morphemes; longer sentences tend to have shorter clauses. Menzerath-Altmann's law appears in many multi-level systems, e.g., proteins, genes, genomes, and genetics (Nikolaou, 2014; Ferrer-I-Cancho & Forns, 2010; Shahzad, Mittenthal, & Caetano-Anollés, 2015; Sun & Caetano-Anollés, 2021).

Lastly, Herdan-Heaps' law (Heaps, 1978; Herdan, 1960, 1964; Egghe, 2007) states that as the size of a text increases, the number of discoverable unique words (i.e., word types) decreases as a function of the text's length. For example, Kornai (2002) verified Herdan-Heaps' law in a corpus of 50 million words, i.e., the distribution the type-token ratio of words did not flatten, indicating that adding more text would lead to more unique words. Brants and Franz (2006) created the first one trillion word corpus of English by aggregating internet webpages and they showed that there were nearly 14 million word types in the corpus – again with no indication that all word types had been discovered.

Of the four linguistic laws, to the best of our knowledge only Herdan-Heaps' law has not yet been studied in animal communication systems.

## 3. Materials and methods

We have identified 19 published studies on linguistic laws and primate vocalizations (see supplementary materials). These include species from different taxa. We identified and extracted information for the following variables (together with the bibliographic metadata): species, linguistic law investigated, evidence (i.e., was the linguistic law supported or not by the data), dependent variable (e.g., units, sequences, bouts), signal investigated (e.g., pant hoots, wobbles), signal modality (e.g., wobble cycles, vocal sequences), definition and criteria of the signal (as per the authors' description), the context of production, sample size, number of subjects, sex, age, social system, size of repertoire reported, and any pertinent or interesting comments or findings.

## 4. Results

The oldest paper that we identified is by McCowan, Doyle, and Hanser (2002), who use comparative measures from information theory to compare the development of vocal repertoires of bottle nose dolphins, squirrel monkeys, and humans. No linguistic laws were tested. Instead, they used the Zipf coefficient to approximate diversity and repertoire complexities. Their findings suggest that in species that are capable of vocal learning, repertoire structure diversity decreases and becomes more organized into adulthood, i.e., less entropic. Thus, they report that chuck calls of squirrel monkeys exhibit similar developmental patterns as in human language acquisition. Other authors use a similar approach in that they investigate power law coefficients in relation to Shannon entropy (Kershenbaum et al., 2021), study developmental trends from infancy to adulthood (Gultekin, Hildebrand, Hammerschmidt, & Hage, 2021), or compare one or more linguistic laws across different species, e.g., macaques, marmosets, and uakaris (Bezerra, Souto, Radford, & Jones, 2011; Ferrer-I-Cancho & Hernández-Fernández, 2013; Ferrer-I-Cancho et al., 2013; Kershenbaum et al., 2021).

The majority of the current published research, however, investigates a linguistic law within a single species' repertoire: chimpanzees (Fedurek, Zuberbühler, & Semple, 2017), geladas (Gustison, Semple, Cancho, & Bergman, 2016), gibbons (Clink, Ahmad, & Klinck, 2020), gorrilas (Watson, Heesen, Hedwig, Robbins, & Townsend, 2020), indris (Zanoli et al., 2020), and macaques (Semple, Hsu, & Agoramoorthy, 2010). Follow-up studies on the same species, but perhaps on different aspects of vocalizations, include (some of) the same authors, e.g., the work on Menzerath's law in geladas (Gustison et al., 2016) or Zipf's law of abbreviation in macaques, marmosets, and uakaris (Semple et al., 2010).

Out of the 19 studies, six investigate Menzerath's law and six investigate Zipf's law of abbreviation. A further three investigate both laws. Regarding Menzerath's law in primate vocal repertoires, there is overwhelming support reported, but with some caveats. Gustison et al. (2016) find that adult male gelada vocalizations follow Menzarath's law. Gustison and Bergman (2017) report that wobble cycle duration was shorter when the number of wobbles or lip smacks was greater. The authors also report that Menzerath's law was identified separately in both inhaled and exhaled wobbles and suggest therefore that the compression of vocal signals by geladas operates at multiple levels (as Menzerath's law does at different linguistic levels in language).

Menzerath's law is also found in tarsiers, titi monkeys, and male Bornean gibbons (Clink & Lau, 2020). These findings raise several important points for consideration when studying linguistic laws in animal communication systems (see also Semple et al. (2021)). First, sexual dimorphism may play an important part in the evolutionary development of vocal repertoires (e.g., in gorillas, who differ greatly in size by sex). Additionally, sexes may differ with regard to the context of production. Thus, ideally the vocalizations of both sexes need to be tested to prevent bias, e.g., Odom, Hall, Riebel, Omland, and Langmore (2014). Second, at what level or which part of the vocal production is analyzed (e.g., duration of notes, proportion of call types, single vs multi-unit sequences) and how a sequence is defined, i.e., which criteria are used to separate the units from each other (e.g., length of silence gaps). Third, a linguistic law may be found in certain call types or certain levels of analysis, but not others. For example, Watson et al. (2020) investigated close-calls of mountain gorillas in sequences. The authors initially found positive evidence for Menzerath's law, but then report that the relationship was due to the difference between single and multi-unit sequences (leaving single units out of the analysis resulted in longer sequences being typically composed of longer units). Hence, Watson et al. (2020) report that close calls by mountain gorillas only partially adhere to Menzerath's law.

Other than mountain gorillas, the only non-human great ape that has so far been studied with regard to linguistic laws is the chimpanzee. Fedurek et al. (2017) find support for Menzerath's law in the number and duration of calls within the pant hoot and for entire vocal sequences. They also report that these findings hold between the duration of adjacent phases in the pant hoot. Pant hoots were investigated in the context of feeding and traveling and only pant hoots produced by males and that contained the climax phase were included in their investigation.

The results from studies of various species and whether their vocalizations adhere to Zipf's law of abbreviation are more mixed. Bezerra et al. (2011) investigate 12 call types of marmosets (excluding predator-specific alarm calls) and 7 from uakaris produced by adults across all contexts. They find no support for Zipf's law of abbreviation. Likewise, Gultekin et al. (2021) find no support for Zipf's law of abbreviation in marmosets throughout their development. However, Semple et al. (2010) examine the full repertoire of macaque vocalizations, in all contexts, from all age ranges and sexes. They show that more frequent vocalizations are shorter in duration. In a follow up study, Semple, Hsu, Agoramoorthy, and Cancho (2013) report that the support for Zipf's law of abbreviation that they found is not an artefact of their previous analysis of mean call duration.

Support for Zipf's law of abbreviation and Menzerath's law was recently reported by Valente et al. (2021) for indris' songs, in contexts of territory defense and long distance communication. Although not an investigation of any particular linguistic law in indris, we note the work of Zanoli et al. (2020), who undertake a Levenshtein distance analysis of adult male and female indri songs. They report that the songs of female indris are less stereotyped that those of males. This work again highlights the importance of evaluating both sexes.

Lastly, we note two studies on whether gibbons' vocalizations adhere to Zipf's law of abbreviation and Menzerath's law. Clink, Tasirin, and Klinck (2020) find no support for Zipf's law of abbreviation, but they report strong support for Menzerath's law. Huang, Ma, Ma, Garber, and Fan (2020) report that both laws are confirmed in male gibbon calls. These two studies highlight the importance of which species and which calls are considered and how sequences are defined.

Clink et al. (2020) study solo singing bouts in multi-phase vocal sequences by male Bornean gibbons. These lengthy singing bouts are comprised of a discrete number of note types with a large repertoire of phrases. Sequences of notes are calculated by two-second breaks or more. The authors also report a strong negative correlation between the number of notes in a phase and the notes' mean duration. They conclude that individual variation produces strong individual signatures. In contrast, the study by Huang et al. (2020) focuses on western black-crested gibbons and Cao-vit gibbons. The authors study the loud morning calls of sub-adult and adult males in various contexts. They define a song bout as all notes in a song with silence periods of less than ten minutes, where a note is a single continuous sound produced either through inhalation or exhalation. They report that the most common notes of the male gibbons follow Zipf's law and that longer sequences follow Menzerath's law.

## 5. Discussion

Here we focus on reported findings of linguistic laws in the vocalization systems of primates because of its potential link to the origin and evolution of speech. However, we note that this review is restricted because it does not include studies on gestures, facial expressions, or multi-modal systems (cf. Liebal, Slocombe, and Waller (2022)). Although such work is beyond the scope of this paper, we have nevertheless begun to collect this information for future research, as we plan to integrate more animal species and communicative signals for meta-analysis.

We find that studies of primate vocalizations and Menzerath's law are in general supportive. However, investigations of Zipf's law of abbreviation in primates are more mixed. Nevertheless, we highlight that within the research of linguistic laws in primate vocalizations, most studies are descriptive in nature. Therefore, we encourage future researchers to consider the four levels of analysis raised by Semple et al. (2021) and what questions they aim to address.

Finally, while we acknowledge the disparate nature of the studies, there are issues and limitations regarding the comparability of species, vocalization types (e.g., duration, number of calls per phase), choice of dependent variables, sexual differences, context of call production, age range, among others. To address these issues, we provide a list in the supplementary materials in which we highlight differences in authors' definitions of units and sequences, the number of subjects in their sample, sample size of vocalizations, and discrepancies in reported repertoire sizes. Moreover, we note that how to test for linguistic laws is still an area of debate even within linguistics, e.g., Zipf's distribution (Piantadosi, 2014). Thus, we hope that studies on animal communication will strengthen this research avenue and ultimately help to elucidate how language evolved in our species.

## 6. Supplementary Materials & Acknowledgements

See: `https://github.com/bambooforest/jcole2022`. Funding from the Swiss NSF: SM & MM (#186841) and KZ, ALF & QG (#185324).

## References

Aitchison, L., Corradi, N., & Latham, P. E. (2016). Zipf's law arises naturally when there are underlying, unobserved variables. *PLOS Computational Biology*, *12*(12), e1005110.

Altmann, G. (1980). Prolegomena to Menzerath's law. *Glottometrika*, *2*(2), 1–10.

Bezerra, B. M., Souto, A. S., Radford, A. N., & Jones, G. (2011). Brevity is not always a virtue in primate communication. *Biology Letters*, *7*(1), 23–25.

Brants, T., & Franz, A. (2006). Web 1T 5-gram Version 1 (2006). *Linguistic Data Consortium, Philadelphia*.

Clink, D. J., Ahmad, A. H., & Klinck, H. (2020). Brevity is not a universal in animal communication: evidence for compression depends on the unit of analysis in small ape vocalizations. *Royal Society Open Science*, *7*(4), 200151.

Clink, D. J., & Lau, A. R. (2020). Adherence to Menzerath's law is the exception (not the rule) in three duetting primate species. *Royal Society Open Science*, *7*(11), 201557.

Clink, D. J., Tasirin, J. S., & Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Current Zoology*, *66*(2), 173–186.

Egghe, L. (2007). Untangling Herdan's law and Heaps' law: Mathematical and informetric arguments. *Journal of the American Society for Information Science and Technology*, *58*(5), 702–709.

Fedurek, P., Zuberbühler, K., & Semple, S. (2017). Trade-offs in the production of animal vocal sequences: insights from the structure of wild chimpanzee pant hoots. *Frontiers in Zoology*, *14*(1), 50.

Ferrer-I-Cancho, R., & Forns, N. (2010). The self-organization of genomes. *Complexity*, *15*(5), 34–36.

Ferrer-I-Cancho, R., & Hernández-Fernández, A. (2013). The failure of the law of brevity in two new world primates. *Glottotheory*, *4*(1), 45–55.

Ferrer-I-Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., & Semple, S. (2013). Compression as a universal principle of animal behavior. *Cognitive Science*, *37*(8), 1565–1578.

Gultekin, Y. B., Hildebrand, D. G. C., Hammerschmidt, K., & Hage, S. R. (2021). High plasticity in marmoset monkey vocal development from infancy to adulthood. *Science Advances*, *7*(27), eabf2938.

Gustison, M. L., & Bergman, T. (2017). Divergent acoustic properties of gelada and baboon vocalizations and their implications for the evolution of human speech. *Journal of Language Evolution*, *2*(1), 20–36.

Gustison, M. L., Semple, S., Cancho, R. Ferrer-i, & Bergman, T. J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, *113*(19), E2750–E2758.

Heaps, H. S. (1978). *Information retrieval, computational and theoretical aspects.* Academic Press.

Herdan, G. (1960). *Type-token mathematics.* Mouton.

Herdan, G. (1964). *Quantitative linguistics.* Butterworth.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*, 89–96.

Huang, M., Ma, H., Ma, C., Garber, P. A., & Fan, P. (2020). Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law. *Animal Behaviour*, *160*, 145-155.

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52.

Kershenbaum, A., Demartsev, V., Gammon, D. E., Geffen, E., Gustison, M. L., Ilany, A., & Lameira, A. R. (2021). Shannon entropy as a robust estimator of Zipf's law in animal vocal communication repertoires. *Methods in Ecology and Evolution*, *12*(3), 553–564.

Kornai, A. (2002). How many words are there? *Glottometrics*, *4*, 61–86.

Levshina, N., & Moran, S. (2021). Efficiency in human languages: Corpus evidence for universal principles. *Linguistics Vanguard*, *7*(S3).

Liebal, K., Slocombe, K. E., & Waller, B. M. (2022). The language void 10 years on: multimodal primate communication research is still uncommon. *Ethology Ecology & Evolution*, *34*(3), 274–287.

McCowan, B., Doyle, L. R., & Hanser, S. F. (2002). Using information theory to assess the diversity, complexity, and development of communicative repertoires. *Journal of Comparative Psychology*, *116*(2), 166–172.

Menzerath, P. (1954). *Die Architektonik des deutschen Wortschatzes.* F. Dümmler.

Nikolaou, C. (2014). Menzerath–Altmann law in mammalian exons reflects the dynamics of gene structure evolution. *Computational Biology and Chemistry*, *53*, 134–143.

Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E., & Langmore, N. E. (2014). Female song is widespread and ancestral in songbirds. *Nature Communications*, *5*(1), 1–6.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin and Review*, *21*(5), 1112–1130.

Semple, S., Cancho, R. Ferrer-i, & Gustison, M. L. (2021). Linguistic laws in biology. *Trends in Ecology & Evolution*, *37*(1), 53–66.

Semple, S., Hsu, M., Agoramoorthy, G., & Cancho, R. Ferrer-i. (2013). The law of brevity in macaque vocal communication is not an artefact of analysing mean call durations∗. *Journal of Quantitative Linguistics*, *20*(3), 209–217.

Semple, S., Hsu, M. J., & Agoramoorthy, G. (2010). Efficiency of coding in macaque vocal communication. *Biology Letters*, *6*(4), 469–471.

Shahzad, K., Mittenthal, J. E., & Caetano-Anollés, G. (2015). The organization of domains in proteins obeys Menzerath-Altmann's law of language. *BMC Systems Biology*, *9*(1), 1–13.

Sun, F., & Caetano-Anollés, G. (2021). Menzerath–Altmann's law of syntax in RNA accretion history. *Life*, *11*(6), 489.

Valente, D., De Gregorio, C., Favaro, L., Friard, O., Miaretsoa, L., Raimondi, T., Ratsimbazafy, J., Torti, V., Zanoli, A., Giacoma, C., & Gamba, M. (2021). Linguistic laws of brevity: conformity in *Indri indri*. *Animal Cognition*, *24*(4), 897–906.

Wacewicz, S., & Żywiczyński, P. (2015). Language evolution: Why Hockett's design features are a non-starter. *Biosemiotics*, *8*(1), 29–46.

Watson, S. K., Heesen, R., Hedwig, D., Robbins, M. M., & Townsend, S. W. (2020). An exploration of Menzerath's law in wild mountain gorilla vocal sequences. *Biology Letters*, *16*(10), 20200380.

Yu, S., Xu, C., & Liu, H. (2018). Zipf's law in 50 languages: its structural pattern, linguistic interpretation, and cognitive motivation. *arXiv preprint arXiv:1807.01855*.

Zanoli, A., De Gregorio, C., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., Giacoma, C., & Gamba, M. (2020). Sexually dimorphic phrase organization in the song of the indris (*Indri indri*). *American Journal of Primatology*, *82*(6), e23132.

Zipf, G. K. (1936). *The psychobiology of language.* Routledge.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley.

# GREAT APE VOCAL REPERTOIRES ARE ALL SIMILAR IN SIZE: NOW WHAT?

Steven Moran*[1,2], Marco Maiolini[1], and Adriano Lameira*[3]

*Corresponding Authors: steven.moran@unine.ch, Adriano.Lameira@warwick.ac.uk
[1]Institute of Biology, University of Neuchâtel, Neuchâtel, Swizterland
[2]Department of Anthropology, University of Miami, Coral Gables, USA
[3]Department of Psychology, University of Warwick, Coventry, UK

There is an uncanny similarity between the size of nonhuman great ape vocal repertoires and that of the phonological inventories documented in human languages. Great ape repertoires vary from a conservative estimate, and lower-bound, of 18 calls in both species of gorillas (*Gorilla*), to the low-to-mid 20s in chimpanzees and bonobos (*Pan*), to the mid 30s in orangutans (*Pongo*). These numbers are in line with the average number of contrastive speech sounds in modern and in ancient and reconstructed languages.

This observation suggests a common evolutionary pressure for a two digit system repertoire, one that existed before the origin of language, and present at the split with our evolutionary ancestors. The difference between systems is well documented – language combines speech sounds (or signs) into an open-ended communication system that *infinitely* produces new words and meanings. And although we do not yet know what led to the cognitive abilities that gave rise to this system in *Homo*, there are many similarities between the vocal repertoires of great apes and those in languages left to explore.

Like language, great ape calls are produced in non-random combinations (Lameira et al., 2021). Like speech, orangutan vocal repertoires are composed of consonant- and vowel-like calls (Lameira, 2014). Orangutan vocalizations also involve modest airstream and articulatory control (Wich et al., 2008) and experiments done in captivity show precise vocal fold motor control, including voicing (Lameira & Shumaker, 2019). And along with other cognitive skills, e.g., tool making, some great ape vocalizations are arguably learned (Hopkins, Taglialatela, & Leavens, 2007) with some groups in captivity reportedly transmitting raspberries to their young, which they use to get the attention of caretakers (Hopkins et al., 2007). Finally like languages, the call repertoires of great apes are shaped by socio-cultural factors (Lameira et al., 2022), they have dialects (Crockford, Herbinger, Vigilant, & Boesch, 2004), change through contact (Mitani & Gros-Louis, 1998), and their existence and diversity are under the threat of extinction (Meijaard et al., 2011).

We conducted an extensive review of the existing literature (66 published articles) and found that call repertoires are similar in size across extant great apes. However, we also discovered that there is little in terms of a comparative articulatory and acoustic analysis of these vocalizations. This is because there is a serious gap in the descriptions of the articulatory features of great apes in the literature, including for both vocal and gestural data.

A step towards a comprehensive comparative phonetic analysis of articulations of all great apes is needed to shed light on the similarities between vocal communication systems and how each evolved in its own right over the last 6-7 million years of divergent evolution. It is crucial for language evolution research that we have articulatory and acoustic phonetic analyses, of the sort conducted by Perlman and Clark (2015) through audio-visual recordings, for all great ape species (and ideally for each community). This would allow, for example, more studies along the lines of Grawunder et al. (2022), who explored the evolution of the vowel space in chimpanzees. The challenge, however, is not trivial because it is difficult to collect the necessary field data – unlike in linguistics, in which fieldworkers have access to speakers and can also ask them questions directly.

So far we have identified the gaps in the literature and created an ontology of the disparate terminology used in great ape studies. Our research suggests that one pertinent area to explore for acoustic analysis is the spectral envelope, so that we can ask whether prosodic features can be modified to give information to the receiver (Zimmermann, Leliveld, & Schehka, 2013). Such findings will shed light on whether paralinguistic features are shared across all great apes. Another area ripe for investigation is what effects exist due to sexual dimorphism, e.g., in orangutans and gorillas, and to understand how to categorize these vocalizations.

We know that the size of great ape vocal repertoires are similar across all extant great apes and that we share eight innate vocalizations, e.g., screaming, crying, laughing (Anikin, Bååth, & Persson, 2018). However, there is still much to learn about what we share vocally, and gesturally (Liebal, Slocombe, & Waller, 2022), and how and when we diverged linguistically.

## Supplementary Materials & Acknowledgements

Supplementary materials and all references are available online.[1] SM & MM were funded by the Swiss NSF (#186841). ARL was funded by the UKRI Future Leaders Fellowship (MR/T04229X/1).

## References

Anikin, A., Bååth, R., & Persson, T. (2018). Human non-linguistic vocal repertoire: Call types and their meaning. *Journal of Nonverbal Behavior*, *42*(1), 53–80.

---

[1] `https://github.com/bambooforest/great_ape_vocal_repertoires`

Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology*, *110*(3), 221–243.

Grawunder, S., Uomini, N., Samuni, L., Bortolato, T., Girard-Buttoz, C., Wittig, R. M., & Crockford, C. (2022). Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200455.

Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, *73*(2), 281–286.

Lameira, A. R. (2014). The forgotten role of consonant-like calls in theories of speech evolution. *Behavioral and Brain Sciences*, *37*(6), 559–560.

Lameira, A. R., Alexandre, A., Gamba, M., Nowak, M. G., Vicente, R., & Wich, S. (2021). Orangutan information broadcast via consonant-like and vowel-like calls breaches mathematical models of linguistic evolution. *Biology Letters*, *17*(9), 20210302.

Lameira, A. R., Santamaría-Bonfil, G., Galeone, D., Gamba, M., Hardus, M. E., Knott, C. D., Morrogh-Bernard, H., Nowak, M. G., Campbell-Smith, G., & Wich, S. A. (2022). Sociality predicts orangutan vocal phenotype. *Nature Ecology & Evolution*, *6*(5), 644–652.

Lameira, A. R., & Shumaker, R. W. (2019). Orangutans show active voicing through a membranophone. *Scientific Reports*, *9*(1), 1–6.

Liebal, K., Slocombe, K. E., & Waller, B. M. (2022). The language void 10 years on: multimodal primate communication research is still uncommon. *Ethology Ecology & Evolution*, *34*(3), 274–287.

Meijaard, E., Buchori, D., Hadiprakarsa, Y., Utami-Atmoko, S. S., Nurcahyo, A., Tjiu, A., Prasetyo, D., Christie, L., Ancrenaz, M., Abadi, F., et al.. (2011). Quantifying killing of orangutans and human-orangutan conflict in Kalimantan, Indonesia. *PLOS ONE*, *6*(11), e27491.

Mitani, J., & Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees: Tests of three hypotheses. *Behaviour*, *135*(8), 1041–1064.

Perlman, M., & Clark, N. (2015). Learned vocal and breathing behavior in an enculturated gorilla. *Animal Cognition*, *18*(5), 1165–1179.

Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2008). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, *50*(1), 56–64.

Zimmermann, E., Leliveld, L., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of emotional communication: From sounds in nonhumman mammals to speech and music in man* (Vol. 116, p. 132). Oxford University Press Oxford, UK.

# ON THE RELATION BETWEEN CONTEXT DEPENDENCY AND VOCABULARY IN HUMAN LANGUAGE AND BIRDSONG

Takashi Morita*[1,2], Ryosuke O. Tachibana[3], Kazuo Okanoya[3,4,5], and Hiroki Koda[2,3]

*Corresponding Author: tmorita@alum.mit.edu
[1]SANKEN, Osaka University, Ibaraki, JAPAN
[2]Primate Research Institute, Kyoto University, Inuyama, JAPAN
[3]Graduate School of Arts and Sciences, the University of Tokyo, Tokyo, JAPAN
[4]Advanced Comprehensive Research Organization, Teikyo University, Tokyo, JAPAN
[5]Behavior and Cognition Joint Research Laboratory, RIKEN Center for Brain Science, Wako, JAPAN

Human language production is often characterized by long dependency on previous output records, or *context*. Recent studies on deep learning-based models of language generation have shown that the dependent context can range over hundreds of words. In this study, we investigated the relation between context dependency and vocabulary type/size, by case-studying English. We found that the long context dependency vanished when words were replaced with their grammatical categories and purely syntactic dependency was considered, which suggests that the long dependency in human language is largely attributed to semantic factors. We also performed clustering in word-embedding spaces and showed that a larger number of clusters (i.e., larger vocabulary) led to longer context dependency. Moreover, a parallel analysis of birdsong (of Bengalese finch, *Lonchura striata* var. *domestica*) revealed the opposite relation between context dependency and vocabulary size (i.e., larger vocabularies shortened dependency length) and moderate vocabulary sizes amounted to comparable dependency lengths to English syntax. We also showed that English phoneme sequences have a much shorter dependency length, which casts doubt on the previous generalization in comparative studies that birdsong is more homologous to human language phonology than to syntax.

## 1. Introduction

Human language production is often characterized by long dependency on previous output records, or *context* (Larson, 2017). Recent studies on deep learning-based models of language generation have shown that the dependent context can range over hundreds of words (Khandelwal, He, Qi, & Jurafsky, 2018; Dai et al., 2019). In this study, we investigated the relation between context dependency and another characteristic property of human language: rich vocabulary. Specifically, we clustered English words into various types/numbers of categories and analyzed its effect on the length of context dependency.

The results of the English analysis will be discussed in comparison with context dependency in birdsong, reproducing our recent study on Bengalese finch (*Lonchura striata* var. *domestica*; Morita, Koda, Okanoya, & Tachibana, 2021).

As pointed out above, the rich vocabulary is one of the most distinguished properties of human language; no other animal seems to handle tens of thousands of vocal categories like human language words. Thus, it is of interest to see whether birdsong exhibits language-like long-distance dependency if it consists of a larger vocabulary, and our previous study addressed this question using the same analytical paradigm as in this study. Here, we complement the cross-specific investigation of the relation between vocabulary size and context dependency through an analysis of English data.

## 2. Materials & Methods

### 2.1. *Measuring Context Dependency by Deep Language Modeling*

We measured context dependency in human language and birdsong by using a deep neural network (Khandelwal et al., 2018; Dai et al., 2019; Morita et al., 2021). We first trained a neural network on the language modeling task: the network computed the predictive probability $P(x_t \mid x_1, \ldots, x_{t-1})$ of each token $x_t$ (e.g., word, phoneme, or birdsong syllable) conditioned on the preceding tokens $x_1, \ldots, x_{t-1}$ (i.e., context). Then, the trained network was used to compute the predictive probability of test tokens conditioned on the full and truncated contexts (Fig. 1A). Intuitively, truncation of a context decreases the predictive probability when dependent tokens are excluded from the context. Thus, the effective context length (ECL) was defined by the minimum length of the truncated context where the difference in the predictive probabilities based on the two contexts faded away; in practice, we adopted the canonical threshold of $1\%$ difference in perplexity following the previous studies (Khandelwal et al., 2018; Dai et al., 2019; Morita et al., 2021).

Following Morita et al. (2021), we adopted a Transformer with six layers, eight heads, hidden dimensionality of 512 (for both the self-attention and feed-forward modules), and relative position encoding (Dai et al., 2019) for language modeling. See Morita et al. (2021, S2 Text) for more details including the training procedure.

### 2.2. *Human Language Data*

We studied the English portion of Wiki40B dataset for word-level analysis of context dependency owing to syntax and/or semantics (Guo, Dai, Vrandečić, & Al-Rfou, 2020). The raw text data were tokenized into word(-like) sequences using the Stanza package of Python (Qi, Zhang, Zhang, Bolton, & Manning, 2020). This resulted in 1,542,787,693 training and 85,462,957 test tokens, which were chunked into 19,873,689 and 1,104,835 paragraphs respectively based on the tags in Wiki40B. Due to limitations in computational resources and time, we performed the analysis of context dependency on only 10,000 test paragraphs (but still amounting to 84,664,199 tokens) that were randomly selected among those

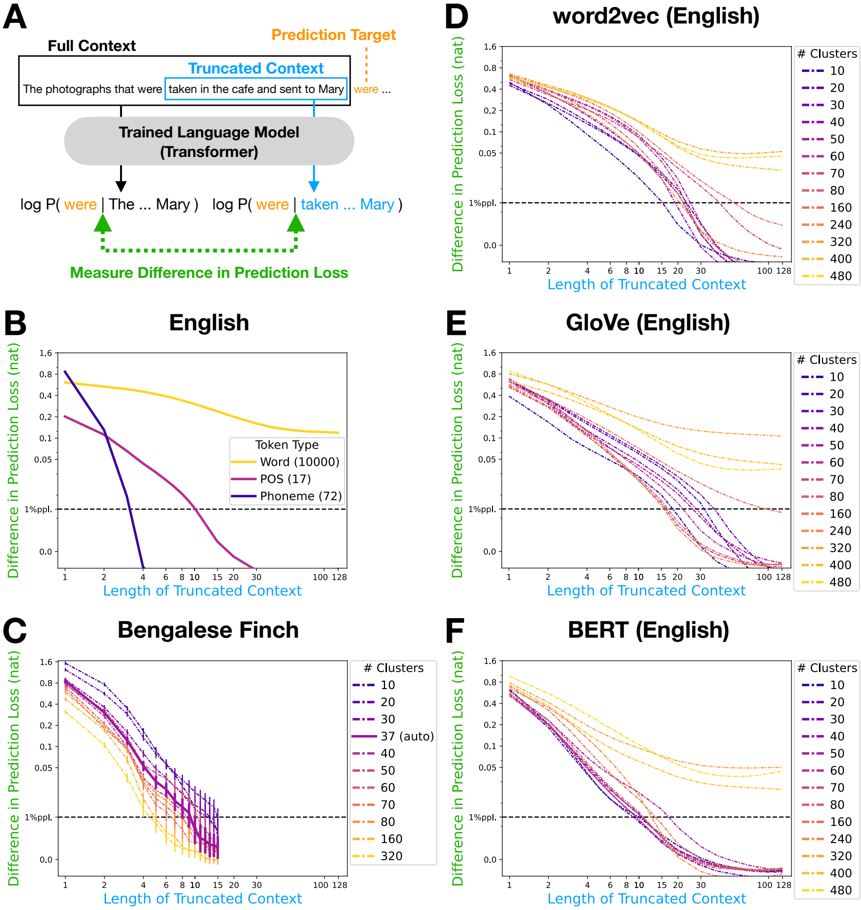


Figure 1. (A) Metric of context dependency based on deep language modeling. Log predictive probabilities of test tokens were computed conditioned on the full context (framed with a black line) and truncated context (framed with a light-blue line). The effective context length (ECL) is the minimum length of the truncated context where the prediction difference went below a threshold (1% in perplexity, or ppl.). (B–F) The differences in the mean loss (negative log probability) between the truncated- and full-context predictions of Wiki40B English words (B), part-of-speech (POS) tags (B), and $k$-means cluster labels in word2vec (D), GloVe (E), and BERT (F) word-embedding spaces; phoneme transcription of LibreSpeech (B); and Bengalese finch songs, consisting of syllable sequences (C). The x-axis corresponds to the length of the truncated context. Both of the axes are in log scale. The horizontal dashed line indicates the canonical threshold at 1% ppl. The error bars in (C) represent the 90% confidence intervals estimated from 10,000 bootstrapped samples, so that the loss difference is statistically significant ($p < 0.05$) if the lower side of the intervals is above the threshold.

containing at least 128 tokens.

Three major types of word representation were investigated:

- 10,000 most frequent word tokens, replacing the rest with a special label `<unk>`.
- Part-of-speech (POS) tags in the universal dependency format (Nivre et al., 2016), induced by a pretrained model provided in Stanza (Qi et al., 2020).
- L2-normalized $k$-means clustering of three pretrained word embeddings ($k := 10, 20, \dots, 80, 160, 320, 400, 480$):
  - Word2vec embeddings (Mikolov, Chen, Corrado, & Dean, 2013).
  - GloVe embeddings (Pennington, Socher, & Manning, 2014)
  - Word embeddings of BERT (Devlin, Chang, Lee, & Toutanova, 2019).

  Only alphanumeric words included in the BERT's vocabulary (the smallest among the three) were clustered and the rest were labeled with `<unk>`.

The POS representation removed semantic distinctions among words (e.g., apple, dog → NOUN; give, run → VERB; a, the → DET) and, thus, allowed us to benchmark a purely syntactic context dependency. Word embeddings are real-valued vector representations that are derived ultimately from co-occurrence statistics of words. Word embeddings are known to reflect syntactic and semantic properties of words in numerical ways. Therefore, clustering in the embedded space allowed us to manipulate the vocabulary size and inquire into its relation with context dependency.

In addition to the word-level analyses above, we also assessed the context dependency in phoneme sequences (i.e., phonotactics). We used the phoneme transcription of the LibriSpeech corpus (Panayotov, Chen, Povey, & Khudanpur, 2015) that was induced by the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017).[1] The Transformer language model was trained on the `train-clean-360` portion of the corpus (consisting of 13,611,485 tokens extracted from 104,008 wav files), and `test-clean` portion (203,760 tokens from 2,620 wav files) was used for estimating the dependency length. We adopted the original wav files of the corpus as the unit of sequence.

## 2.3. *Birdsong Data*

We reproduced the estimation of context dependency in Bengalese finch song, originally reported in (Morita et al., 2021). In that study, we performed Transformer language modeling on syllable sequences produced by eighteen birds. Syllables were classified into discrete categories by an unsupervised clustering method with speaker normalization (see also Morita & Koda, 2020) and the resulting symbolic representation was used for the language modeling and analysis of context dependency. The number of syllable categories was automatically esti-

---

[1] Precomputed transcription is provided by Lugosch, Ravanelli, Ignoto, Tomar, and Bengio (2019) in `https://doi.org/10.5281/zenodo.2619474`.

mated by the clustering method as 37. We also reported the context dependency based on finer- and more coarse-grained classifications of the syllables (into 10, 20, . . . . 80, 160, 320 categories) by L2-normalized $k$-means clustering.[2] Here, we reproduced the analysis of context dependency for each of those syllable clusterings. The dataset consisted of 457,992 and 6,557 syllables (segmented into 7,779 and 100 sequences by recording sessions) for training and test respectively.

## 3. Results

While sequences of raw English words exhibited the context dependency beyond 128 tokens, the effective dependency length dropped to eleven when the words were replaced with their POS tag (Fig. 1B). The dependency in phoneme sequences was even shorter—only on four tokens—despite their larger vocabulary size than the POS categories.

Manipulation of the vocabulary size by word-embedding clustering showed that larger vocabularies led to longer context dependency (Fig. 1D–F). By contrast, the context dependency of Bengalese finch song decreased as finer-grained syllable classifications were adopted (Fig. 1C; due to the small size of test data, the dependency length was estimated considering the statistical significance of the difference between full- and truncated-context predictions against the threshold with $p < 0.05$, following Morita et al., 2021). When coded by the automatically detected vocabulary consisting of 37 syllable categories, the birdsong exhibited effective dependency on eight tokens. Smaller vocabularies with 10–30 categories increased the dependency length to ten to fifteen tokens. Conversely, larger vocabularies decreased the dependency length, up to five when 160/320 syllable categories were assumed.

## 4. Discussion

We found that the word-level context dependency of English became drastically shorter when tokens were replaced with their POS tag. This result suggests that the long context dependency in human language is mostly attributed to semantic factors and purely syntactic dependencies can be handled relatively locally, by referring to eleven recent tokens in the production history. This view is also consistent with the positive correlation between the number of word-embedding clusters and the dependency length; finer-grained clusterings of embeddings recovered more semantic information in the original continuous space, which led to the longer dependency length in turn.

By contrast, finer-grained classification of Bengalese finch syllables decreased the dependency length (as originally reported in Morita et al., 2021). This find-

---

[2] All the clustering results are publicly available in `https://doi.org/10.17605/OSF.IO/R6PAQ`.

ing indicates that expanding vocabulary size does not lead to language-like long-distance dependency by itself. Specifically, minor acoustic variations in Bengalese finch syllables—which are encoded in fine-grained classifications but ignored in coarse-grained classifications—are unlikely to carry semantics-like information (Okanoya, 2007; Berwick, Okanoya, Beckers, & Bolhuis, 2011; Miyagawa, Berwick, & Okanoya, 2013). Instead, acoustic properties of the birdsong syllables are known to be affected by surrounding syllables (Wohlgemuth, Sober, & Brainard, 2010) and, thus, paying more attention to local context improved the prediction of fine-grained syllable categories (e.g., five previous tokens when 160 and 320 categories are assumed). Such local interactions are more akin to human language phonotactics; as demonstrated by our analysis of phoneme sequences in English, phonotactic dependencies are much shorter (four tokens; although longer and potentially unbounded phonotactic dependencies have been suggested for limited patterns in several languages; Sapir & Hoijer, 1967) than syntactic dependencies encoded by POS sequences (eleven tokens).

Meanwhile, it should be noted that Bengalese finch song—under the assumption of a moderate number of syllable categories (10–40)—exhibited longer dependency on eight to fifteen tokens than English phonotactics (four tokens). The estimated dependency length of the birdsong instead amounted to that of English syntax (eleven tokens), which casts doubt on the previous generalization in comparative studies that birdsong is more homologous to human language phonology than to syntax (Berwick et al., 2011). The non-local context dependency comparable to human language syntax also implies that the birdsong cannot be modeled efficiently by traditional $n$-gram grammars (Hosino & Okanoya, 2000) because exponentially more transitional rules are needed as the dependency length increases. The long dependencies would be captured more succinctly by latent structures as in hidden Markov models (Rabiner, 1989; Katahira, Suzuki, Okanoya, & Okada, 2011) and hierarchical grammars (Berwick, 2015; Morita & Koda, 2019), or by distributed representation as in biological/artificial neural networks (Nishikawa, Okada, & Okanoya, 2008; Dai et al., 2019).

## Acknowledgements

This study was supported by JSPS Grant-in-aid for Scientific Research on Innovative Areas #4903 (Evolinguistic; JP17H06380) to HK and KO, JSPS Grant-in-Aid for Scientific Research (JP19KT0023, JP21H03781) to ROT, and for Early-Career Scientists (JP21K17805) to TM, the JST Core Research for Evolutional Science and Technology 17941861 (JPMJCR17A4) to HK and ACT-X 21454934 (JPMJAX21AN) to TM, and the Mitsubishi Foundation Research Grants in the Natural Sciences (202111014) to TM. We also gratefully acknowledge the support of the Academic Center for Computing and Media Studies, Kyoto University, regarding the use of their supercomputer system.

## References

Berwick, R. C. (2015). Mind the gap. In Á. J. Gallego & D. Ott (Eds.), *50 years later: Reflections on Chomsky's aspects* (pp. 1–12). Cambridge, MA: MITWPL.

Berwick, R. C., Okanoya, K., Beckers, G. J., & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Science*, *15*(3), 113–121.

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q., & Salakhutdinov, R. (2019). Transformer-XL: Attentive language models beyond a fixed-length context. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 2978–2988). Florence, Italy.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 4171–4186). Minneapolis, Minnesota.

Guo, M., Dai, Z., Vrandečić, D., & Al-Rfou, R. (2020). Wiki-40B: Multilingual language model dataset. In *Proceedings of the 12th language resources and evaluation conference* (pp. 2440–2452). Marseille, France.

Hosino, T., & Okanoya, K. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in bengalese finches. *Neuroreport*, *11*(10), 2091–2095.

Katahira, K., Suzuki, K., Okanoya, K., & Okada, M. (2011). Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PLOS ONE*, *6*(9), 1–9.

Khandelwal, U., He, H., Qi, P., & Jurafsky, D. (2018). Sharp nearby, fuzzy far away: How neural language models use context. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 284–294). Melbourne, Australia.

Larson, B. (2017). Long distance dependencies. *Oxford Bibliographies*.

Lugosch, L., Ravanelli, M., Ignoto, P., Tomar, V. S., & Bengio, Y. (2019). Speech model pre-training for end-to-end spoken language understanding. In *Proceedings of INTERSPEECH* (pp. 814–818). Graz, Austria.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. In *Proceedings of INTERSPEECH* (pp. 498–502). Stockholm, Sweden.

Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Y. Bengio & Y. LeCun (Eds.), *Proceedings of the 1st international conference on learning representations (ICLR)*. Scottsdale, Arizona.

Miyagawa, S., Berwick, R., & Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology*, *4*, 71.

Morita, T., & Koda, H. (2019). Superregular grammars do not provide additional explanatory power but allow for a compact analysis of animal song. *Royal Society Open Science*, *6*(7), 190139.

Morita, T., & Koda, H. (2020). Exploring TTS without T using biologically/psychologically motivated neural network modules (ZeroSpeech 2020). In *Proceedings of interspeech 2020* (pp. 4856–4860). Shanghai, China.

Morita, T., Koda, H., Okanoya, K., & Tachibana, R. O. (2021). Measuring context dependency in birdsong using artificial neural networks. *PLOS Computational Biology*, *17*(12), e1009707.

Nishikawa, J., Okada, M., & Okanoya, K. (2008). Population coding of song element sequence in the Bengalese finch hvc. *European Journal of Neuroscience*, *27*(12), 3273–3283.

Nivre, J., Marneffe, M.-C. de, Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., & Zeman, D. (2016). Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the tenth international conference on language resources and evaluation (LREC'16)* (pp. 1659–1666). Portorož, Slovenia.

Okanoya, K. (2007). Language evolution and an emergent property. *Current Opinion in Neurobiology*, *17*(2), 271–276.

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). Librispeech: An ASR corpus based on public domain audio books. In *2015 ieee international conference on acoustics, speech and signal processing (ICASSP)* (pp. 5206–5210). South Brisbane, Queensland, Australia.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Doha, Qatar.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th annual meeting of the association for computational linguistics: System demonstrations* (pp. 101–108). Online.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Sapir, E., & Hoijer, H. (1967). *The phonology and morphology of the Navaho language.* Berkeley, CA: University of California Press.

Wohlgemuth, M. J., Sober, S. J., & Brainard, M. S. (2010). Linked control of syllable sequence and phonology in birdsong. *Journal of Neuroscience*, *30*(39), 12936–12949.

# THE EFFECT OF NETWORK STRUCTURE ON LEXICAL VARIABILITY IN SIGN LANGUAGE COMMUNITIES

Katie Mudd[1], Marnix Van Soom[1], Connie de Vos[2], and Bart de Boer[1]

*katie.mudd@ai.vub.ac.be
[1]AI lab, Vrije Universiteit Brussel, Belgium
[2]Department of Communication and Cognition, Tilburg University, the Netherlands

The study of sign languages allows us to observe what features characterize a language in its early stages. The initial phase of emergence is characterized by a high degree of lexical variability, where synonyms for a concept appear to coexist in a community. Overtime, lexical variation may persist; for example, it was found that Kata Kolok, a sign language used in a small community in a Balinese village with a tightly knit social network, exhibits a high degree of lexical variability (Mudd et al., 2020). Meanwhile, in communities with different social configurations, we suggest that lexical variation may decrease within groups of frequent interlocutors (i.e. on the *local* level) but remain high across the entire population (i.e. on the *global* level). For example, British Sign Language (BSL), used by the large Deaf community across the UK, was found to be lexically uniform at the local level (i.e. within regions), but at the global level BSL actually has a higher degree of lexical variability than Kata Kolok (Mudd, Lutzenberger, Schembri, Ohanin, & Stamp, 2022).

Several factors have been proposed to explain the retention of lexical variability in sign language emergence: de Vos (2011) suggests that populations that retain a high degree of lexical variability typically have frequent face-to-face interaction, are small and lack formal deaf education. Previous models have investigated population size (Thompson, Raviv, & Kirby, 2020) and the role of shared context (Mudd, de Vos, & de Boer, 2022), but the role of *network structure* on lexical variation has yet to be explored. In addition, variation at the local and global levels has yet to be considered, with previous models implicitly considering variation only at one, unspecified level.

Building on the shared context model by Mudd et al. (2022), which formalizes the hypothesis that social and psychological information shared within groups of signers affords the use of iconic signs and thus the retention of lexical variability, we investigate if network structure can explain differences in the degree of lexical variation at the local and global levels. In more detail, in the initial phase of the model, agents improvise signs to refer to different concepts based on their social and psychological information, such that agents in the same groups are likely to produce similar signs. However, we made the arguably unrealistic assumption that group membership does not influence interaction patterns. Here, we generalize our previous model by allowing the interaction probability between pairs of agents to depend on their group membership through Newman (2003)'s assortativity coefficient $r$. A continuous spectrum of *assortative network structures* can thus be probed by varying $r$, from agents interacting completely at random ($r = 0$) to interacting only with others in the same group ($r = 1$).

To study the effect of assortative network structure on lexical variability at the local (i.e. within group) and global (i.e. across the entire population) levels, we keep a fixed set of parameters[1] and vary the value of $r$, as shown in Fig 1. We observe that $r$ correlates with the lexical variability at both levels. Interestingly, the local and global levels decouple only at high values of $r$, i.e. for strongly assortative networks.

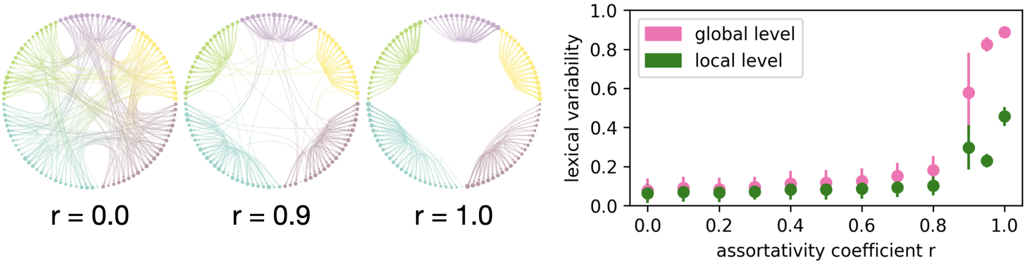


Figure 1. Network visualizations (left) depict interactions (edges) between groups (nodes colored according to group membership) for different values of $r$. The plot (right) shows the effect of assortative network structure on local and global lexical variability. For high values of $r$, the degree of variation at the global level is much higher than at the local level.

Social networks tend to high values of $r$ (Newman, 2003), and if we suppose that the interaction between BSL signers living in different regions can be modeled by a strongly assortative network structure, our result may help explain the observed difference between local and global levels. Network structure is likely one suitable candidate in determining the degree of local and global lexical variability in languages, most pronounced for communities consisting of several locally tightly knit groups (where $r$ is high). Finally, with this model, a range of assortative network configurations can be considered, without needing to compare different categorizations of sign languages (Hou & Vos, 2021).

---

[1] The fixed model parameters are: number of agents = 100, number of groups agents are randomly assigned to = 5, number of concepts = 10, number of bits for the form and culturally salient features = 10, overlap between the culturally salient features and the form = 0.9. The model is run for 2000 time steps. 100 iterations of the run are averaged, shown as mean ± standard deviation in Fig 1.

## References

de Vos, C. (2011). Kata Kolok Color Terms and the Emergence of Lexical Signs in Rural Signing Communities. *The Senses and Society*, *6*(1), 68–76.

Hou, L., & Vos, C. de. (2021). Classifications and typologies: Labeling sign languages and signing communities. *Journal of Sociolinguistics*, 1–8.

Mudd, K., de Vos, C., & de Boer, B. (2022). Shared context facilitates lexical variation in sign language emergence. *Languages*, *7*(1), 31.

Mudd, K., Lutzenberger, H., Schembri, A., Ohanin, O., & Stamp, R. (2022). The social structure of signing communities and lexical variation: A cross-linguistic comparison of three unrelated sign languages. In *Theoretical Issues in Sign Language Research (TISLR).* Osaka, Japan.

Mudd, K., Lutzenberger, H., Vos, C. de, Fikkert, P., Crasborn, O., & Boer, B. de. (2020). The effect of sociolinguistic factors on variation in the Kata Kolok lexicon. *Asia-Pacific Language Variation*, *6*(1), 53–88.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*, 167–256.

Thompson, B., Raviv, L., & Kirby, S. (2020). Complexity can be maintained in small populations: a model of lexical variability in emerging sign languages. In *Proceedings of the 13th International Conference on the Evolution of Language* (p. 3). Brussels, Belgium.

# MODELING THE ORIGINS OF COMMUNICATION: ALIGNMENT OF INTERESTS AND SOCIAL CUES THEREOF

ROLAND MÜHLENBERND*[1] and ANDREAS BAUMANN[2]

*Corresponding Author: muehlenbernd@leibniz-zas.de
[1]Leibniz-Centre General Linguistics, Berlin, Germany
[2]University of Vienna, Vienna, Austria

Alignment of interests is a fundamental factor for the evolutionary stability of communication. We will apply tools from evolutionary game theory to analyze the relationship between the degree of alignment of interests among interlocutors and the evolutionary stability of their potential signaling strategies. We will study different scenarios and show that there is an evolutionary benefit for adopting a communication strategy which is conditioned on social cues: hints about the degree of alignment of interests in a given situation. We will show that this strategy i) is the only evolutionarily stable strategy in a scenario with social cues, ii) has very low underlying requirements (with respect to alignment of interests) to be evolutionarily stable, and iii) is less reliable for the receiver, since it determines the sender to be partially dishonest.

## 1. Introduction

A fundamental underlying factor for the stability of communication is *alignment of interests* (AOI). Searcy and Nowicki (2005) illustrate in a number of scenarios that the evolutionary stability of animal signaling depends on the AOI between sender and receiver. They distinguish a) scenarios where interests overlap, such as food begging signals or alarm calls, where stable signaling is suitable and common, from b) scenarios where interests diverge or oppose, such as mating signals or displays of aggression, where stable signaling is rather unexpected and can only be explained through additional conditions. Naturally, AOI does also play an important role in human language. For example, the Gricean *cooperation principle* tells us that interlocutors' interests are generally aligned in at least one aspect, namely to communicate appropriately and cooperatively (Grice, 1975).

Often, the purpose of a communicative signal depends on the extent to which interests are aligned. For example, birds such as the great titmouse use alarm calls in cases when there is actual danger (high AOI), but they also use them as 'false alarm' in situations where they want to move other birds away from some food source (low AOI) (cf. Matsuoka, 1980). That is, titmice are able to adapt their signaling strategies not only towards the actual information they want to transfer (the source of danger) but also towards properties of the receiver, i.e., social information. This form of deception is frequently found in human language as well. For example, the usage of polite speech can serve for sweetening the conversation (high AOI) or for persuading the hearer for doing something that is rather in the interest of the speaker (low AOI).

In this article, we will apply evolutionary game theory to show how the stability of communication strategies under population dynamics depends on the AOI in the underlying scenario. Moreover, we will demonstrate that the establishment of stable communication is more likely when senders adapt their signaling strategies towards *social cues*, i.e., hints about the AOI between interlocutors in a given situation. Finally, we will show that the benefit of the integration of social cues into communication comes with a hitch: it promotes dishonest signaling.

## 2. The mating game: communication with (partially) aligned interests

We study a fundamental question concerning the origin of communication: how can a signal become an entity of (meaningful) communication at all? We take a look at a simple form of communication, namely *signaling*, where a sender produces a signal that evokes a response from a receiver. One important aspect of signaling being sustainably successful is the alignment of interests between a sender and a receiver. When the sender has an advantage in evoking a particular response from a receiver by sending a signal, and the receiver, in turn, has an advantage in performing this response upon the signal, then both parties have an aligned interest in successful communication.

To emphasize the role of alignment of interests in communication, we want to draw the attention to a signaling scenario where the interests between sender and receiver are only partially aligned. Diverse game-theoretic versions of this scenario have been studied with tools from evolutionary game theory (cf. Crawford & Sobel, 1982; Grafen, 1990).We will present a simple model version of this scenario, which we call the *mating game*.

The scenario is as follows: we have a situation where a male and a female of a particular species get together to get involved in a potential mating act. The male can be a high quality ($H$) or low quality ($L$) type. The female can agree to mate ($M$) or not to mate ($\bar{M}$). For any type of male, high or low type, it is always evolutionarily beneficial to mate with the female. For the female, however, it is evolutionarily beneficial to mate with high type males and reject low type males. The simplest form of the underlying payoff structure of this scenario is given in Figure 1(a), where each strategy profile is attributed to two utility values, the first for the male (row player), the second for the female (column player). Utility values are either 1 (evolutionarily beneficial) or 0 (evolutionarily detrimental).

In this scenario, interests are only aligned for the case where the male is a high type. It can be shown that sustainable signaling cannot evolve. Let's assume that high type males start sending a signal to the females, upon which females mate, otherwise they don't mate. That works fine for the females as long as only high type males send that signal. But now low type males have an evolutionary benefit

| | $M$ | $\bar{M}$ |
|---|---|---|
| $H$ | $1, 1$ | $0, 0$ |
| $L$ | $1, 0$ | $0, 1$ |

(a) underlying payoff structure

| | $\langle M, \bar{M} \rangle$ | $\langle \bar{M}, \bar{M} \rangle$ |
|---|---|---|
| $\langle s, s \rangle$ | $1, 1/2$ | $0, 1/2$ |
| $\langle s, \bar{s} \rangle$ | $1/2, 1$ | $0, 1/2$ |
| $\langle \bar{s}, s \rangle$ | $1/2, 0$ | $0, 1/2$ |
| $\langle \bar{s}, \bar{s} \rangle$ | $0, 1/2$ | $0, 1/2$ |

(b) mating game

Figure 1. The mating game's (a) underlying payoff and (b) normal-form representation.

in sending the signal, and will therefore adopt it. But then the signal is meaningless for females, because it lost its function, namely to distinguish high type from low type males. Therefore, they can ignore it and signaling stops working.

This line of thought can be studied mathematically. Whether a signaling strategy is sustainable can be mathematically deduced by proofing if it is an evolutionarily stable strategy (Maynard Smith & Price, 1973). As a first step, we define a space of signaling strategies. We define a sender strategy as a tuple $\langle x_H, x_L \rangle$, whereby $x_H$ represents the signaling behavior of high types, and $x_L$ represents the signaling behavior of low types. In its simplest version, our scenario entails four different sender strategies: $\langle s, s \rangle$, $\langle s, \bar{s} \rangle$, $\langle \bar{s}, s \rangle$ and $\langle \bar{s}, \bar{s} \rangle$, whereby $s$ stands for sending a signal, $\bar{s}$ stands for not sending a signal. Similarly, a receiver strategy is defined as a tuple $\langle y_s, y_{\bar{s}} \rangle$, whereby $y_s$ represents the response behavior upon receiving a signal, and $y_{\bar{s}}$ the response behavior upon not receiving a signals. For simplicity, we assume that receivers never mate upon not receiving a signal. Then the receiver strategies are: $\langle M, \bar{M} \rangle$ and $\langle \bar{M}, \bar{M} \rangle$.

From the underlying payoff structure we can deduce a utility table over pairs of signaling strategies by computing *expected utilities* (see the Online Appendix Sec. B) which represent how beneficial (on average) a combination of a sender strategy and a receiver strategy (a *strategy profile*) is for the sender (first value) and the receiver (second value). The resulting utility table is given in Figure 1(b).

From these utilities we can identify strategy profiles that are evolutionarily stable (Maynard Smith & Price, 1973). Selten (1980) has shown for asymmetric games, such as the mating game, that a strategy profile is evolutionarily stable if and only if the row player's utility is unique maximum in its column, and the column player's utility is a unique maximum in its row. This means that any unilateral switch out of an evolutionarily stable strategy profile leads to a strictly lower payoff for the switching player. The readers can check for themselves that none of the strategy profiles of Figure 1(b) is evolutionarily stable. Particularly, the strategy profile that represents separating signaling ($\langle s, \bar{s} \rangle$, $\langle M, \bar{M} \rangle$) is not evolutionarily stable, since the sender has a benefit from switching to $\langle s, s \rangle$. And then, ($\langle s, s \rangle$, $\langle M, \bar{M} \rangle$) is not evolutionarily stable, since the receiver is not strictly worse off when switching to $\langle \bar{M}, \bar{M} \rangle$. A similar argument can be found for any other strategy profile of the mating game in Figure 1(b).

| | $M$ | $\bar{M}$ |
|---|---|---|
| $H$ | $1, 1$ | $0, 0$ |
| $L$ | $0, 0$ | $1, 1$ |

(a) underlying payoff structure

| | $\langle M, \bar{M} \rangle$ | $\langle \bar{M}, \bar{M} \rangle$ |
|---|---|---|
| $\langle s, s \rangle$ | $1/2, 1/2$ | $1/2, 1/2$ |
| $\langle s, \bar{s} \rangle$ | $1, 1$ | $1/2, 1/2$ |
| $\langle \bar{s}, s \rangle$ | $0, 0$ | $1/2, 1/2$ |
| $\langle \bar{s}, \bar{s} \rangle$ | $1/2, 1/2$ | $1/2, 1/2$ |

(b) lethal mating game

Figure 2. Lethal mating game's (a) underlying payoff and (b) normal-form representation.

Let us now consider a scenario where interests are completely aligned, as represented by the payoff structure in Figure 2(a). This could represent a mating scenario where the female is extremely selective and aggressive. For example, after agreeing upon mating, the female tests the strength of the male before the actual mating act happens. High-quality males pass the test and mate with the female, whereas low-quality males fail the test and get killed by the female before the mating can happen. Here, low types have an evolutionary benefit of not initiating a mating act with this type of female. Let us call this game the lethal mating game. Its normal-form game representation is given in Figure 2(b).

This game has exactly one evolutionarily stable strategy profile, namely $(\langle s, \bar{s} \rangle, \langle M, \bar{M} \rangle)$, since any unilateral switch of strategy leads to a strictly worse payoff for the switching player. In this strategy profile, high-quality males send a signal upon which females mate, and low-quality males don't send a signal and won't get killed by the female. The bottom line of this section is this: in contrary to the standard mating game, signaling in the lethal mating game is evolutionarily stable, particularly because the interests of sender and receiver are totally aligned.

## 3. Transfer to a scenario of cultural evolution

Note that these mating game scenarios study stability aspects of signaling strategies in the light of biological evolution, where utility represents fitness which drives the rate of biological reproduction. But when it comes to the evolution of human communications and language, we are frequently concerned with cultural evolution, where utility equals fitness which drives the rate of behavioral reproduction, for example in form of imitation or learning. Since the concept of evolutionary stability is assumed to be applicable for biological evolution as well as cultural evolution (O'Connor, 2020), we can be agnostic about the concrete underlying dynamics. Moreover, very often we can transfer games that describe scenarios that a relevant in the light of biological evolution to scenarios that are relevant in the light of cultural evolution. We will exemplify this hereafter.

Here is an idea how we can reinterpret the mating game as a communication scenario primarily exposed to dynamics of cultural evolution: Assume that we have a group of individuals which happen to use alarm calls for dangerous events, such as raptor attacks, etc. Let's say that the individuals can distinguish between events of high danger ($H$) and events of low danger ($L$). In our game, the row player is a watchpost, who is in charge of giving alarm calls. The column player is a gatherer, who is doing a utility-relevant activity, such as picking berries.

In a first scenario the gatherer is selfish: he is picking the berries for himself. He has an interest to be warned whenever there is a high danger event, so he can stop picking and look for a save place to move to ($M$). However, in a low danger event, it is more beneficial not to move ($\bar{M}$) and continue picking berries. The watchpost, however, has an interest in always interrupting the gatherer, since there will be more berries left for her when her watchpost shift is over. The underlying payoff structure of this scenario is exactly the one in Figure 1(a), and the resulting communication game is exactly the one in Figure 1(b). And the analysis leads to the same conclusion: the strategy profile of separating signaling ($\langle s, \bar{s}\rangle$, $\langle M, \bar{M}\rangle$) is not evolutionarily stable. And also the reason is the same: interests are not aligned. Picking a lot of berries is good for the gatherer, but bad for the watchpost.

Let us think about a similar scenario where the gatherer is not picking berries for himself, but for the whole group instead, including the watchpost. Then the watchpost would prefer only to interrupt the gatherer when the gatherer's life is at stake (in high danger situations), but otherwise prefers him to continue picking berries, since more work done by the gatherer means more food for the whole group, including the watchpost herself. This scenario is exactly the one presented in Figure 2(a). Accordingly, Figure 2(b) represents the resulting communication game. Here again, the conclusion is the same: the separating signaling strategy ($\langle s, \bar{s}\rangle$, $\langle M, \bar{M}\rangle$) is evolutionarily stable since interests are aligned. Picking a lot of berries is good for both the gatherer and for the watchpost.

## 4. Combined communication games and social cues

Note that both scenarios of the last section differ with respect to the social behavior of the gatherer. In the first case, he is selfish; in the second case, he is social. In this section we want to merge both scenarios to a combined one, where the gatherer is sometimes selfish and sometimes social. In more general terms, in such a *combined scenario* the situation of totally aligned interests (Figure 2(a)) occurs with a frequency $p$ in $(0, 1)$, and the situation of partially aligned interests (Figure 1(a)) occurs otherwise, with frequency $1 - p$. When we combine the utility tables of both situations weighted by probability $p$, we obtain the underlying payoff structure as given in Figure 3(a). Its normal form game representation with signaling options is given in Figure 3(b), which we label *communication game 1*.

Here, it can be shown that the separating signaling strategy ($\langle s, \bar{s}\rangle$, $\langle M, \bar{M}\rangle$) is evolutionarily stable if and only if $p > 1/2$; for the proof see the Online Appendix Sec. A.1. In other words, when situations of totally aligned interests occur more frequently than situations of partially aligned interests, then the separating signaling strategy is evolutionarily stable. With respect to our concrete example, this is the case when gatherers behave more often social than selfish.

|  | $M$ | $\bar{M}$ |
|---|---|---|
| $H$ | $1, 1$ | $0, 0$ |
| $L$ | $1-p, 0$ | $p, 1$ |

(a) underlying payoff structure

|  | $\langle M, \bar{M}\rangle$ | $\langle \bar{M}, \bar{M}\rangle$ |
|---|---|---|
| $\langle s, s\rangle$ | $1 - {}^{p}/_{2}, {}^{1}/_{2}$ | ${}^{p}/_{2}, {}^{1}/_{2}$ |
| $\langle s, \bar{s}\rangle$ | ${}^{1}/_{2} + {}^{p}/_{2}, 1$ | ${}^{p}/_{2}, {}^{1}/_{2}$ |
| $\langle \bar{s}, s\rangle$ | ${}^{1}/_{2} - {}^{p}/_{2}, 0$ | ${}^{p}/_{2}, {}^{1}/_{2}$ |
| $\langle \bar{s}, \bar{s}\rangle$ | ${}^{p}/_{2}, {}^{1}/_{2}$ | ${}^{p}/_{2}, {}^{1}/_{2}$ |

(b) communication game 1

Figure 3. The (a) underlying payoff and (b) normal-form representation of communication game 1. The alignment of interests between row and column player depends on $p$.

|  | $\langle M, \bar{M}\rangle$ | $\langle \bar{M}, \bar{M}\rangle$ |
|---|---|---|
| $\langle s, s, s, s\rangle$ | $1 - \frac{p}{2}, \frac{1}{2}$ | $\frac{p}{2}, \frac{1}{2}$ |
| $\langle s, s, \bar{s}, s\rangle$ | $\boxed{1, \frac{1+p}{2}}$ | $\frac{p}{2}, \frac{1}{2}$ |
| $\langle s, s, \bar{s}, \bar{s}\rangle$ | $\frac{1+p}{2}, 1$ | $\frac{p}{2}, \frac{1}{2}$ |
| $\langle \bar{s}, \bar{s}, \bar{s}, \bar{s}\rangle$ | $\frac{p}{2}, \frac{1}{2}$ | $\frac{p}{2}, \frac{1}{2}$ |

Figure 4. A subset of four selected strategies in communication game 2. The full game entails the same finding: there is only one evolutionarily stable strategy pair (and only for $p > 0$), namely $(\langle s, s, \bar{s}, s\rangle, \langle M, \bar{M}\rangle)$ (the boxed entry).

The definition of communication game 1 implies that individuals cannot distinguish situations with totally aligned interests from those with partially aligned interests. Let's assume that the sender is able to distinguish both scenarios, for example via *social cues* that give hints about the receiver's attitude. For example, when the sender can identify if the gatherer is behaving selfish or social, she can condition her strategy accordingly. We can model this by extending the sender's strategy space to a quadruple $\langle x_{Ht}, x_{Hp}, x_{Lt}, x_{Lp}, \rangle$, where $x_{Ht}$ is the signaling behavior of a sender in a scenario with totally aligned interests in a high-danger event, $x_{Hp}$ is the signaling behavior of a sender in a scenario with partially aligned interests in a high-danger event, and so on. In each case, the sender can either send a signal $s$ or not $\bar{s}$, leading to $2^4 = 16$ sender strategies. We call the resulting game *communication game 2*. Figure 4 shows a subset of strategies in communication game 2. The full utility table is given in the Online Appendix Sec. A.2.

It can be shown that there is exactly one evolutionarily stable strategy pair, namely $(\langle s, s, \bar{s}, s\rangle, \langle M, \bar{M}\rangle)$, where the sender always sends the signal in high-danger events, but also in low-danger events in case that the receiver is selfish. Importantly, this strategy pair is evolutionarily stable for any $p$ in $(0, 1)$; for the proof see the Online Appendix Sec. A.2. The bottom line of this section is this: there is a much higher AOI requirement for evolutionary stability *without a sender's access to social cues* (communication game 1; requirement: $p > \frac{1}{2}$) than *with a sender's access to social cues* (communication game 2; any $p$ in $(0, 1)$). This is our main finding and we will come back to this point in the conclusion.

Note that the scenarios in this section assume that signaling comes without costs. In the Online Appendix Sec. C, we studied communication games 1 and 2 with signaling costs. The result shows that signaling costs strengthen the conditions for separating signaling strategies to be evolutionarily stable in both games. Since the separating signaling strategy in communication game 2 is not conditioned on social cues, we can conclude that signaling costs weaken the prerequisites for the evolutionary stability of signaling strategies that involve social cues.

## 5. Reliability and honesty

We want to study the communication strategies with respect to the concepts *reliability* and *honesty*. Following Searcy and Nowicki (2005), a signal is reliable if i) some characteristic of the signal (including its presence/absence) is consistently correlated with some attribute of the signaler or its environment, and ii) receivers benefit from having information about this attribute. Moreover, a signal does not have to be perfectly reliable, but it is enough when the signal is *honest on average*, such that the receiver on average is better off assessing the signal than ignoring it.

Following these lines of thought, we define a signal $s$ to be honest, if and only if it is sent in situations such that the expected response of the receiver is evolutionarily beneficial to the receiver. Hereof, we define the reliability of a signal $s$ as the ratio of the honest usage of $s$ to the total usage of $s$. More precisely, when a sender uses a signal $s$ with frequency $f_T$ in total and uses it honestly with frequency $f_H$, then the reliability of the signal is given by the ratio $\frac{f_H}{f_T}$. Thus, reliability of a signal is a value between 0 and 1 (never/always honestly used).

Let's take a look at the reliability of signal $s$ in sender strategy $\langle s, s, \bar{s}, s \rangle$ as part of the only stable signaling strategy pair $(\langle s, s, \bar{s}, s \rangle, \langle M, \bar{M} \rangle)$ in communication game 2. Here, totally aligned situations occur with frequency $p$ and partially aligned situations with frequency $1 - p$. High danger and low danger events are implicitly assumed to occur both with frequency $\frac{1}{2}$. This gives us the following frequencies for the four situations that the sender can distinguish: $f(x_{Ht}) = \frac{p}{2}$, $f(x_{Lt}) = \frac{p}{2}$, $f(x_{Hp}) = \frac{1-p}{2}$ and $f(x_{Lp}) = \frac{1-p}{2}$. In the stable signaling strategy $\langle s, s, \bar{s}, s \rangle$ the sender sends the signal with total frequency $f_T = f(x_{Ht}) + f(x_{Hp}) + f(x_{Lp}) = \frac{2-p}{2}$. Furthermore, since the signal $s$ is only honestly sent when there is a high danger event, we get frequency $f_H = f(x_{Ht}) + f(x_{Hp}) = \frac{1}{2}$. Now we can compute the reliability of signal $s$ as follows: $\frac{f_H}{f_T} = \frac{\frac{1}{2}}{\frac{2-p}{2}} = \frac{1}{2-p}$.

From this analysis we can follow that the reliability of the signal $s$ in strategy $\langle s, s, \bar{s}, s \rangle$ is greater than $\frac{1}{2}$ for any $p$ in $(0, 1)$. However, this result also shows that strategy $\langle s, s, \bar{s}, s \rangle$ is not totally reliable (save for the borderline case with $p = 1$). On the other hand, it is easy to see that in the evolutionarily stable sender strategy $\langle s, \bar{s} \rangle$ of communication game 1, signal $s$ has a reliability value of 1 (since $s$ is only and always sent in $H$), therefore it is totally reliable.

## 6. Conclusion

We have argued that classical game-theoretical models of animal signaling can be transferred to scenarios that might have played an important role in the evolution of human language. The evolutionary analyses of the models in this paper have demonstrated a well-known general aspect of any form of communication: the evolutionary stability of meaningful signaling/communication depends to a great degree on the alignment of interests (AOI) between sender and receiver. The more particular results of our analyses are connected to the juxtaposition of the communication games 1 and 2. This analysis showed that access to social cues lowers the underlying AOI requirements for signaling strategies being evolutionarily stable.

We conclude that the ability to integrate social cues into communication is a conducive feature for stable communication and therefore a potentially important propulsive factor in the evolution of language. The latter analysis showed that the integration of social cues comes with a hitch, though: it promotes dishonest signaling. But then, dishonest signaling isn't necessarily something adverse in the presented scenario. It can also be seen as a form of punishment that the sender exerts on the receiver for selfish behavior. Here, dishonesty might have evolved for a good cause, namely to suppress the behavior of nonprosocial individuals in order to enhance sociality. This aspect ought to be pursued in subsequent research.

**Online Appendix**: The Online Appendix is part of the proceeding's supplementary material and accessible under https://doi.org/10.6084/m9.figshare.20004395

**Funding**: Roland Mühlenbernd was funded by the German Research Foundation (DFG) within the SFB 1412 under DFG agreement number 41659133.

## References

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*, 1431–1451.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*, 517–546.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Edited by P. Cole/J. Morgan, New York: Academic Press.

Matsuoka, S. (1980). Pseudo warning call in titmice. *Tori*, *29*, 87-90.

Maynard Smith, J., & Price, G. (1973). The logic of animal conflict. *Nature*(146), 15–18.

O'Connor, C. (2020). *Games in the philosophy of biology.* Cambridge University Press.

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication.* Princeton University Press.

Selten, R. (1980). A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology*, *84*, 93–101.

# MACHINE IMPOSTORS IMPEDE HUMAN DETECTION AND THE SUCCESS OF STABLE CONVENTIONS BY IMITATING PAST INTERACTIONS

THOMAS F. MÜLLER*[1], LEVIN BRINKMANN[1], JAMES WINTERS[2], and NICCOLÒ PESCETELLI[3]

*Corresponding Author: mueller@mpib-berlin.mpg.de
[1]Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany
[2]School of Collective Intelligence, Mohammed VI Polytechnic University, Ben Guerir, Morocco
[3]Department of Humanities and Social Sciences, New Jersey Institute of Technology, Newark, United States

How can successful communication arise and stabilize between humans? Artificial language games have shown that we create novel *conventions* even from minimal communicative means (Galantucci, 2005; Garrod et al., 2007; Scott-Phillips et al., 2009). A convention is defined as the arbitrary solution to a repeated coordination problem (Lewis, 1969), established through, e.g., precedence (Clark, 1996). Meanwhile, results from the perceptual crossing paradigm (Auvray et al., 2009; Barone et al., 2020) have shown that human interaction in minimal environments is distinct from interaction with automated agents in its *reciprocal* patterns, i.e. interdependence between participants. However, these studies deliberately avoid conventional paths to success. In the current study, we build on these experimental approaches to investigate conventional and reciprocal routes to communication in a minimal environment. In a Turing test setup (Turing, 1950) stripped of the use of natural language, we manipulate bot behavior to assess how it affects the detection of bots, and participants' behaviors. Our main hypothesis is that access to past interactions makes bot impostors more deceptive by interrupting convention formation.

We recruited 200 participants online for the experiment and assigned them to one condition in pairs. Their goal was to find out whether their random partner in each trial was the human partner or the bot impostor. Participants interacted within a 2D space containing an orange square and a blue circle. Their only way to communicate was to move the square, while the human partner or the bot moved the circle. In both conditions, bot behaviors were identical replays of previous behavior from human pairs, but their source differed: While bots in the *partner impostor condition* repeated behaviors shown earlier by the participant's own human partner, bots in the *foreign impostor condition* imitated an unrelated participant from a previous partner impostor condition. Participants received feedback on their own and their partner's performance after each trial.

We tested our preregistered predictions via model comparison of mixed-effects models. Predicting performance by condition revealed that participants were more successful in the foreign impostor condition than in the partner impostor condition ($\beta = -1.92$, $SE = 0.16$, $\Delta\text{AIC} = 69$; Fig. 1a). We measured the conventionality of participants' behaviors by the Earth Mover's distance to their own spatial positions over trial blocks. Human performance in the partner impostor condition suffered from conventional behavior, while the foreign impostor condition profited from it ($\beta = 0.30$, $SE = 0.08$, $\Delta\text{AIC} = 10$; Fig. 1b). Last, we measured reciprocity by computing the transfer entropy between participants' and their partners' movements, and found that it was only useful to detect bots in the partner impostor condition ($\beta = 0.17$, $SE = 0.07$, $\Delta\text{AIC} = 3.4$).

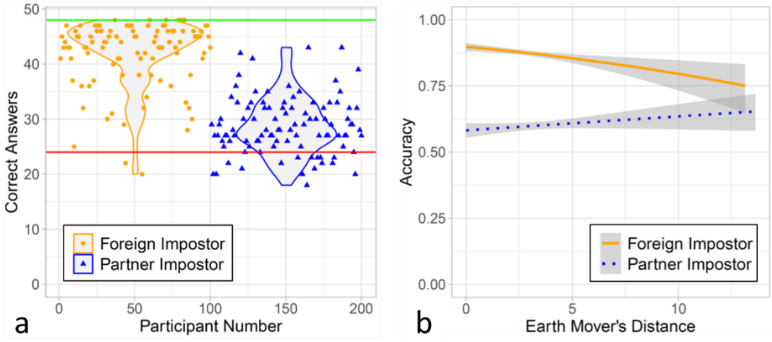


Figure 1. a) Performance results by participant. Violin plots show the density in the two conditions. The top line represents the performance ceiling, the middle line performance at chance. b) Relation between conventionality and accuracy in the two conditions. Note that higher Earth Mover's distance translates to lower conventionality. The shaded area shows the 95% confidence interval.

Our results suggest bots can avoid human detection and prevent conventional behaviors from succeeding by imitating past interactions, emphasizing the role of the interaction history for human communication. Although both conventional and reciprocal behaviors were adaptive under the right conditions, participants struggled when conventionality was maladaptive. We show how manipulating bot behavior can provide new insights into emergent communication, combining ideas from language evolution, pragmatics, and social cognition.

## References

Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27*(1), 32–47. https://doi.org/10.1016/j.newideapsych.2007.12.002

Barone, P., Bedia, M. G., & Gomila, A. (2020). A Minimal Turing Test: Reciprocal Sensorimotor Contingencies for Interaction Detection. *Frontiers in Human Neuroscience*, *14*, 102. https://doi.org/10.3389/fnhum.2020.00102

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*(5), 737–767. https://doi.org/10.1207/s15516709cog0000_34

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–987. https://doi.org/10.1080/03640210701703659

Lewis, D. (1969). *Convention*. Harvard University Press.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*(2), 226–233. https://doi.org/10.1016/j.cognition.2009.08.009

Turing, A. (1950). Computing Machinery and Intelligence. *Mind, 59,* 433–460.

# THE ROLE OF CROSS-CULTURAL DIFFERENCES IN REFERENTIAL STRATEGY EMERGENCE: DATA FROM ARTIFICIAL SIGN LANGUAGE LEARNING

Danielle Naegeli*[1], David Peeters[1], Emiel Krahmer[1], Marieke Schouwstra[2], Yasamin Motamedi[3], and Connie de Vos[1]

*Corresponding Author: d.s.u.naegeli@tilburguniversity.edu
[1]Department of Communication and Cognition, Tilburg University, Tilburg, Netherlands
[2]Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, Netherlands
[3]Centre for Language Evolution, University of Edinburgh, Edinburgh, United Kingdom

The grammatical use of space for referential strategies such as spatial modification and role shift can be found across many sign languages (e.g., Engberg-Pedersen, 1993; Liddell, 2003). Signers, for example, use pointing signs towards arbitrary locations in space to refer to different agents. However, not all sign languages use spatial strategies for non-spatial relations. The rural sign languages Kata Kolok (used in the North of Bali) and Al-Sayyid Bedouin Sign Language (used in southern Israel) have not developed strategies relying on anaphoric use of space (de Vos, 2012; Padden et al., 2010). This raises the question of why there are such typological differences. A proposed explanation attributes the absence of such strategies to the relatively young age of these sign languages, suggesting that these strategies have not developed *yet* (Sandler et al., 2005). However, while Kata Kolok and Al-Sayyid Bedouin Sign Language are younger than, for example, American or British sign language they differ considerably in age and furthermore, another “young” sign language, Nicaraguan Sign Language, developed spatial strategies early on in its emergence (Senghas et al., 1997). Moreover, hearing English-speaking participants in a silent gesture experiment that simulated intergenerational transmission with an iterated learning design produced referential strategies using space anaphorically within just five generations (Motamedi et al., 2021). Therefore, the age of respective sign languages alone cannot serve as an explanation, and hence, the question remains: What can account for such differences across sign languages?

This preregistered study (https://osf.io/w4sgx) aims to investigate another possible factor contributing to spatial strategies emerging in a sign language: To what degree can these typological differences be explained by differences in their respective surrounding co-speech gesture system? Indeed, the Balinese co-speech gesture system is different from most Western ones, for example, using a geocentric instead of an egocentric pointing system (Wassman & Dasen, 2006) – a feature it shares with Kata Kolok (de Vos, 2012). Thus, if the ambient co-speech gesture system is relevant to the development of spatial referential strategies, we should see speakers of Balinese produce strategies that differ from the ones employed by speakers of Indo-European languages like Dutch in a silent gesture task. This in turn could shed light on why Kata Kolok has not developed anaphoric use of space.

We are currently conducting replication studies of Motamedi et al.'s (2021) silent gesture experiment in the Netherlands and in Bali allowing for cross-cultural comparison of the results. Fifty hearing participants of either language group (speakers of Dutch or speakers of Balinese) are paired up and asked to perform a communicative director-matcher task using only gestures. The design of the experimental stimuli is such that participants have to distinguish agents in order for their partner to be able to choose the right option from an array. The experimental set-up was slightly adapted to use the online video conferencing app Zoom to function in a remote setting during the pandemic. Participants' gestured responses are recorded and adhering to an iterated learning design (like Motamedi et al., 2021), the videos of one participant were used in training trials for the following dyad. Using a director-matcher task in combination with iterated learning allows us to analyse the strategies emerging through language use and to record the possible entrenchment of these strategies through the simulation of repeated generational transmission.

Our set-up not only provides the opportunity to explore and compare the strategies produced by participants from different cultural backgrounds but also collects data from individuals that are users of the exact co-speech gesture system that surrounds Kata Kolok signers.

Pilot data from our Balinese experiment has already given us some indications that the strategies developed by participants there do not make use of space the same way as our Dutch participants do and the English-speaking participants in Motamedi et al. (2021) did. Thus, at the conference, we will present our comparative data, discuss what strategies the different participant groups have produced, and evaluate what this suggests about the role of the co-speech gesture system in the emergence of sign languages.

## References

de Vos, C. (2012). Sign-spatiality in Kata Kolok: How a village sign language in Bali inscribes its signing space. PhD Thesis, Max Planck Institute for Psycholinguistics, Nijmegen.

Engberg-Pedersen, E. (1993). *Space in Danish Sign Language. The Semantics and Morphosyntax of the Use of Space in Visual Language.* Hamburg: Signum.

Liddell, S. K. (2003). *Grammar, Gesture and Meaning in American Sign Language.* Cambridge: Cambridge University Press.

Motamedi, Y., Kenny, S., Schouwstra, M., Culbertson, J., & Kirby, S. (2021). The emergence of systematic argument distinctions in artificial sign languages. *Journal of Language Evolution*, 1-22.

Padden C., Meir, I., Aronoff, M., Sandler, W. (2010). The Grammar of Space in Two New Sign Languages. Sign Languages: A Cambridge Language Survey.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences*, *102*(7), 2661-2665.

Senghas, A., Coppola, M., Newport, E. L., & Supalla, T. (1997). Argument Structure in Nicaraguan Sign Language: The Emergence of Grammatical Devices. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Boston University Conference on Language Development 21 Proceedings,* 550-561. Somerville, MA: Cascadilla Press.

Wassmann, J., & Dasen, P. R. (2006). How to Orient Yourself in Balinese Space: Combining Ethnographic and Psychological Methods for the Study of Cognitive Processes. In J. Straub, D. Weidemann, C. Köbl, and B. Zielke (Eds.), *Pursuit of Meaning. Advances in Cultural and Cross-Cultural Psychology* (pp. 351-376). Bielefeld: Transcript Publishers.

# EVOLUTION OF THE WORD ORDER RULES THROUGH COMMUNICATION AND CULTURAL TRANSMISSION: AN ITERATED LEARNING EXPERIMENT

Seiya Nakata*[1,2,4], Akari Ohira[1], and Masanori Takezawa*[1,3,4]

*Corresponding Author: snakata@eis.hokudai.ac.jp, m.takezawa@let.hokudai.ac.jp
[1]Department Behavioral Science, Hokkaido University, Sapporo, Japan
[2]Japan Society for the Promotion of Science, Japan
[3]Center for Experimental Research in Social Sciences, Hokkaido University, Japan
[4]Center for Human Nature, Artificial Intelligence and Neuroscience, Hokkaido University, Japan

The grammatical structure found in human language may have emerged through communication and cultural transmission. To focus on the role of communication in the evolution of word order rules, we designed an iterated learning experiment in which participants use a given set of words. The results of the experiment showed that communication facilitates the cultural evolution of word order rules. In cultural transmission chains with communication, the structural language with word order rules emerged rapidly and remained stable throughout the generations. On the other hand, in the cultural transmission chain without communication, the evolution of word order rules was very slow, and the word order distribution didn't show stability. These findings suggest that word order rules culturally evolve by sharing the intention to communicate.

## 1. Introduction

Human languages have a systematic grammatical structure. For example, the order of subject-object-verb (SOV) is determined according to the grammatical rules of each language (Schouwstra, & de Swart, 2014). The systematic ordering of elements by grammatical rules allows for smooth communication among speakers of the same language. The systematic word order rules, rather than simply listing words, are rare in the communication systems of non-human animals (Hauser, Chomsky, & Fitch, 2002; but see also Suzuki, 2014; Suzuki, Wheatcroft, & Griesser, 2017). The question of where the systematic structure found in human language emerges from is one of the central questions in evolutionary linguistics. Studies using iterative learning experiments have shown that the linguistic structure emerges from a trade-off between learnability (cultural transmission) and expressivity (communication) (Kirby et al., 2008; 2015; Winters, 2015). In these experimental studies, reusable parts of strings corresponding to individual meanings evolved from the random strings of the initial language. They used small and simple meaning spaces and showed the emergence of basic compositional structure. Saldana et al. (2019) expanded the semantic space and investigated the emergence of the compositional hierarchical structure of language in the iterated learning experiments. They made the task more complex and implemented communication among participants. Through a series of iterated learning experiments, they found that compositional linguistic structures having hierarchical word order rules inherent to natural languages evolved under cultural transmission and communication.

In the current study, we focused on the cultural evolution of grammatical structures exhibited as word order rules. Initial language in Saldana et al. (2019) was holistic strings each corresponding to a unique stimulus. They found the emergence of two levels of the hierarchy of constituents: word-like forms and sentence-like structures. In our experiment, the initial language was a sequence of randomly ordered words. Each word referred to either an actor or an action. As there was no systematic rule determining word order, a sequence is regarded as holistic language each of which corresponds to a unique meaning. By employing a similar experimental setting as experiment 2 in Saldana et al. (2019), we investigated whether grammatical rules for determining word orders emerge through the processes of cultural transmission and communication.

We were particularly interested in whether communication plays an essential role in the evolution of the word order rules. We compared a solo condition, in which languages were transmitted in a single transmission chain, with a pair condition, in which pairs of individuals communicate with languages. If communication plays an essential role, we will observe the rapid emergence of a word order rule in the pair condition. If the cultural transmission is sufficient for the emergence of word order rules, we won’t observe much difference between the conditions.

## 2. Methods

### *2.1. Participants*

120 (51 females and 69 males) undergraduates of Hokkaido University, whose native language was Japanese, were recruited. Eighty and forty participants were assigned to the pair condition and the solo condition respectively. Mean age was 19.93 years (*SD* = 1.55; ranging from18 to 23). Participants received 500 JPY

(about 4.33 USD) and additional payment based on the number of successes in the task (described later).

### *2.2. Stimuli*

Participants were asked to learn a combination of a two-second animation and a sentence. All the animation consisted of two actors and the action of each actor (Fig. 1). A sentence was made of four words; two words represent actors (a red rabbit, a blue rabbit, a red bird, or a blue bird) and the other two words represent an action of actors (a rabbit punching, a rabbit getting angry, a bird punching, or a bird getting angry). Each word was randomly made of two to three Japanese letters and participants could not modify the words. Participants were allowed to see the correspondence table of the words and their meaning during the experiment. The initial sentences presented to the participants in the first generation were created by randomly ordering the four words representing the elements in an animation. The order of words representing actions was fixed for representing the chronological order of the actions – the action word appearing first represents the action of the left actor, who first initiates an action, in the animation.

Figure 1. An example of animation and a sentence. Each animation started with an action of a left actor followed by an action of a right actor (the text in brackets was not shown).

### *2.3. Experimental procedures*

The experiment consists of two phases. In the learning phase, participants were asked to memorize pairs of animations and sentences presented on a computer screen. There were 16 pairs of stimuli and eight were randomly selected for the learning phase. In each round of the learning phase, a pair of a sentence and animation was presented on the computer screen for 6 seconds. At the end of each round of presentation, one of two confirmation tests was randomly administered. In the sentence production test, participants were presented with only the animation again and were asked to reproduce the presented sentence by rearranging the four given words. In the animation selection test, participants were presented with only the sentence again and asked to select the corresponding animation from a given set of four animations. Each stimulus was presented three times and the learning phase consisted of 24 rounds (three blocks of eight stimuli).

Next, participants in the pair condition engaged in a communication phase where two participants worked on the communication task for a total of 64 rounds (four blocks of sixteen stimuli). In the communication phase, a total of 16 stimuli were used, including 8 stimuli that did not appear in the learning phase. The 16 stimuli were randomly ordered and presented in each block. The roles of speaker and hearer were alternated every round so that a participant played the roles of a speaker and a hearer twice for each stimulus. Participant assigned a speaker’s role was shown an animation and asked to produce a corresponding sentence by rearranging the four words presented. Participants assigned a hearer’s role were presented with a sentence produced by the speaker and were asked to choose a corresponding animation from four candidate animations. If the animation selected by the hearer matched the one presented to the speaker, both participants received 20 JPY as an additional payment. After each round, both participants received feedback on their communication; the animation presented to the speaker, the sentence produced by the speaker, the animation selected by the hearer were displayed with the cumulative number of successes. After completing the experiment, participants answered a post-questionnaire and received payments based on the number of successes.

Instead of the communication phase, participants in the solo condition engaged in the test phase. Participants were given one of the two tests similar to the ones used in the communication phase of the pair condition – either creating a sentence corresponding to the displayed animation by rearranging the four presented words or choosing the animation corresponding to the displayed sentence out of the four candidates. The number of rounds of the test phase was identical to the communication phase in the pair condition. At the end of every round in the test phase, participants received feedback and 40 JPY was added to the payment for each correct answer.

Participants were organized into an independent transmission chain each with five generations. In the solo condition, 16 pairs of animations and sentences produced by a participant in the third and last blocks of the test phase were transmitted. In the pair condition, one of the participants was randomly selected and the 16 pairs of sentences and animations produced as a speaker in the third and last block of the communication phase were transmitted. In both conditions, eight pairs were randomly selected and displayed in the learning phase of the next generation. The animation was created by Adobe Animate 2021, and the experiment was programmed in oTree v. 3.3.11 (Chen, Schonger, & Wickens, 2016).

## 3. Results

### *3.1. Emergence of word order rules*

In the following, "N1" refers to a word meaning the actor who initiates the action, and "N2" refers to a word meaning the actor reacted to the action of N1. "V1" and "V2" refer to the action of N1 and N2, respectively. All the sentences are made by ordering these four words (N1, N2, V1, and V2). Initial sentences were made by randomly ordering words although V1 always preceded V2 for representing the order of actions. The number of possible orderings is 24. For analyzing the emergence of word order rules, we calculated the Shannon entropy (Shannon, 1948) for a frequency distribution of each order on 16 sentences transmitted to the next generation. If all the 16 sentences were made by a single ordering, the value of entropy becomes zero.

Fig. 2 (left) shows the mean Shannon entropy per generation for each condition. Values in generation zero were calculated based on the initial 8 sentences participants in the first generation learned in the learning phase. In the pair condition, entropy approached zero at the first generation and remained unchanged. In the solo condition, entropy did not change until the third generation and gradually lowered from the fourth generation. It is suggested that a single word order rule quickly emerged in the pair condition while a word order rule very slowly evolved in the solo condition.

We fitted a linear mixed model where conditions, generations, and interaction effect were entered as predicting factor variables (chain was treated as a random effect variable). We then calculated estimated marginal means for post-hoc contrasts. We found that, except for the generation zero, the entropy in the pair condition was significantly smaller than the solo condition in all the generations ($M_{(\text{pair} - \text{solo})} = -2.41 \sim -1.06$, *p*s = .000). In the pair condition, the entropy in the generation zero was significantly larger than the other generations ($M_{(0 - 1, 2, 3, 4, \text{ or } 5)} = 2.18 \sim 2.51$, *p*s = .000). In the solo condition, entropy in the generation 4 and 5 were significantly smaller than the other generations ($M_{(0,1,2 \text{ or } 3 - 4 \text{ or } 5)}$ 0.59~1.38, *p*s = .05 ~ .001).

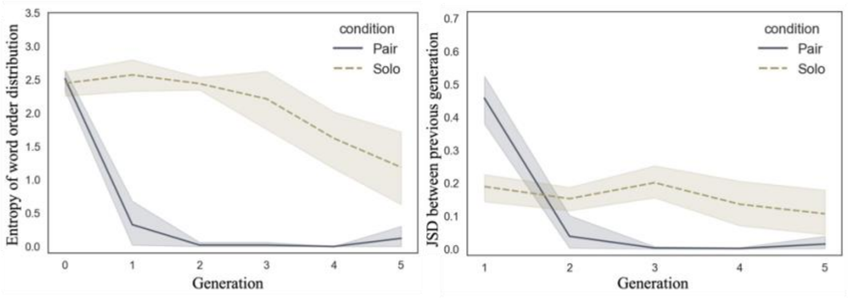


Figure 2. (left) Entropy of word order distribution plotted by generation. Lines and shadows show mean and bootstrapped 95% CI respectively. (right) Similarity of the word order distribution between neighboring generations. Distances between t–1 and t are shown as t on the horizontal axis. Lines and shadows show mean JSD between previous generation and bootstrapped 95% CI respectively.

### *3.2. Word order stability*

For analyzing the similarity of word orders within a chain, we calculated the Jensen–Shannon divergence metric (JSD) as the distance of word order distributions between neighboring generations (i.e., $t$–1 vs. $t$). The lower JSD indicates word order distributions are more similar (Fig. 2 (right)). As we've already discussed, in the pair condition, a single word order rule quickly emerged in generation 1 and remained unchanged for the rest of the generations. This pattern is also exhibited in the trend of JSD distance in the pair condition. On the other hand, in the solo condition, the language changed a little each generation, and no stable word order distribution emerged.

We fitted the similar LMM on the JSD and found that, in the solo condition, JSD were not significantly different across generations ($M$ = –0.05~0.09, $p$s = .082~.999). In the pair condition, JSD in the first generation significantly larger than the other generations ($M_{(1-2\sim5)}$ = 0.42~0.46, $p$s = .000).

### *3.3. Evolved word order rules reflected causal order in the referent*

Table 1 shows the most frequently used word orders in the final generation of each chain. In both conditions, "N1-V1-N2-V2" and "V1-N1-V2-N2" (italics in the table) were the most frequently used although word orders were more diverse in the solo condition. Although any word order could achieve successful communication in the pair condition, we found that the word orders reflecting the order of action in the animation were more likely to evolve in the experiment, which was similar to that of the participants' native language, Japanese.

Table 1. The most frequently used word orders utilized in the final generation (numbers in parentheses are frequencies).

| | *Pair condition* | *Solo condition* |
|---|---|---|
| Chain 1 | *V1-N1-V2-N2* (75%) | *N1-V1-N2-V2* (38%), V1-N1-N2-V2 (38%) |
| Chain 2 | *N1-V1-N2-V2* (100%) | *N1-V1-N2-V2* (100%) |
| Chain 3 | *V1-N1-V2-N2* (100%) | N1-N2-V1-V2 (25%), N1-V1-V2-N2 (25%), V1-N1-N2-V2 (25%), *V1-N1-V2-N2* (25%) |
| Chain 4 | *N1-V1-N2-V2* (100%) | *N1-V1-N2-V2* (50%) |
| Chain 5 | *N1-V1-N2-V2* (100%) | *N1-V1-N2-V2* (100%) |
| Chain 6 | *N1-V1-N2-V2* (97%) | N1-V1-V2-N2 (94%) |
| Chain 7 | *N1-V1-N2-V2* (100%) | *N1-V1-N2-V2* (31%), N1-V1-V2-N2 (31%) |
| Chain 8 | *N1-V1-N2-V2* (97%) | *V1-N1-V2-N2* (50%) |

## 4. Discussion

We investigated the hypothesis that communication facilitates the cultural evolution of the word order rules. Our iterated learning experiments showed that communication is a powerful pressure on the evolution of word order rules. In the pair condition where communication is necessary, structural languages with word order rules emerged rapidly. The word order distributions in the pair condition showed stability throughout the generation. In contrast, in the solo condition, the evolution of word order rules was much slower, and the word order distribution changed from generation to generation.

Specific word orders became dominant in the final generation of all chains in the pair condition. On the other hand, in the solo condition, various word orders remained even in the final generation. Throughout the experiment, participants frequently produced word orders to reflect the order of the actions in the animation. In our experimental design, participants could communicate even with the initial random word order sentences because they could see the correspondence table of the words and their meanings. Nonetheless, it is interesting to note that specific word order rules were preferred. In other words, the emergence of word order rules could be explained by the intent of the communication, not by the need to understand the meanings of sentences. It is suggested that the rapid emergence of word order rules in the pair condition is due to the shared intention to communicate appropriate meaning to each other. However, the impact of an incentive to reproduce the initial language in the solo condition (which would not be present in the pair condition) should also be considered in the future. Finally, the participants' acquired languages could attribute to the high frequency of causal order emergence. Therefore, it would be necessary to conduct future experiments with participants from various language families at various stages of language development.

## Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17H06383. We would like to thank Koji Fujita on the helpful advice for the experimental design.

## References

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88-97.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681-10686.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87-102.

Saldana, C., Kirby, S., Truswell, R., & Smith, K. (2019). Compositional hierarchical structure evolves through cultural transmission: an experimental study. *Journal of Language Evolution, 4*(2), 83-107.

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition*, 131(3), 431-6.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423.

Suzuki, T. N. (2014). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Animal Behaviour*, *87*, 59-65.

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature communications*, *7*(1), 1-7.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(3), 415-449.

# CHALLENGING A WIDELY-ACCEPTED ACCOUNT OF VOWEL METATHESIS IN NAGOYA JAPANESE WITH NO REFERENCE TO PRECEDENCE

KUNIYA NASUKAWA[*1], and SHIN-ICHI TANAKA[2]

[*]Corresponding Author: nasukawa@mail.tohoku-gakuin.ac.jp
[1]Department of English, Tohoku Gakuin University, Sendai, Japan
[2]Department of Language and Information Sciences, University of Tokyo, Japan

To understand the nature of the language faculty as part of the program of research into the origins of human language, it is essential to establish which properties of human language make it species-specific. The Minimalist Program of Generative Grammar sees internal language as a computational system (CS) which generates an infinite number of hierarchically-organised (syntactic) objects, each of which underlies an array of instructions to two interfaces: the Sensory-Motor (SM) systems and the Conceptual-Intentional (CI) systems. In this model of the language faculty, a linguistic expression is constructed by taking lexical items from the lexicon and merging them to form a new object (e.g., $\gamma + \{\alpha, \beta\} \rightarrow \{\gamma\{\alpha, \beta\}\}$). The repeated application of Merge to its own derivatives generates an infinite number of recursively structured expressions, where precedence relations are not encoded between constituents (Hauser, Chomsky & Fitch, 2002; Chomsky, 2010). In this model, precedence relations are viewed as epiphenomenal, a by-product of linearisation taken at the SM interface.

To make all types of structure-building systematically coherent, Precedence-free Phonology (PfP: Backley & Nasukawa, 2020, and others) claims that CS takes not only morpho-syntactic objects but also phonological features called *elements* as its arguments. Then, through the merging of elements CS builds a hierarchical phonological representation for each morpheme before being stored in the lexicon. In this model, as in syntax, precedence is the natural result of computing and interpreting the head-dependent relations which hold between units in a structure. This paper challenges the common assumption that the analysis of phonological processes relies on precedence. It illustrates this by focusing on the following type of vowel metathesis in Nagoya Japanese, which is typically analysed as a swapping process operating between two adjacent vowels *ai*, *ae*, *oi* and *ui*: the palatality of the second vowel becomes an on-glide to the first vowel, while the quality of the first vowel is preserved in the newly formed CV (see also Tanaka, 2022, for an analysis of vowel fusion in Tokyo Japanese).

Table 1. On-gliding of the second member of a sequence of two vowels (‘→’ = ‘corresponds to’).

| | *Tokyo Standard J.* | | *Nagoya J.* |
|---|---|---|---|
| a. | *um**ai*** ‘delicious’ | → | *um**ja:*** |
| | *om**ae*** ‘you, dear’ | → | *om**ja:*** |
| b. | *sug**oi*** ‘amazing, great’ | → | *sug**jo:*** |
| | *zur**ui*** ‘go home’ | → | *zur**ju:*** |

Since PfP makes no reference to precedence relations, it rejects any analysis based on metathesis. Instead, it regards the process in question as coalescence, in which two sounds merge into one by combining their properties (‘resonance’ elements (|I| (dip), |U| (rump), |A|(mass), see Backley, 2011), as illustrated below.

Table 2. Element fusion and |I| salience (salient elements underlined).

| | $V_1$ | | $V_2$ | | $V_{1,2}$ | Traditionally described as | phonetically realised as |
|---|---|---|---|---|---|---|---|
| a. | \|A\| | + | \|I\| | → | \|AI̲\| | *jaː* | [æː] |
| | \|A\| | + | \|AI\| | → | \|AI̲\| | *jaː* | [æː] |
| b. | \|AU\| | + | \|I\| | → | \|AUI̲\| | *joː* | [øː] |
| | \|U\| | + | \|I\| | → | \|UI̲\| | *juː* | [yː] |

Fusion takes place between $V_1$ and $V_2$, which in PfP are structured hierarchically ($V_1$ is dominated by $V_2$) rather than ordered sequentially (Backley, 2021). In addition, a language-specific rule makes |I| (palatality) structurally dependent, meaning that it makes a bigger contribution to the acoustic signal of the fused structure and is therefore perceived as having greater prominence than other elements (THE PRINCIPLE OF PHONETIC REALISATION OF HEAD-DEPENDENCY STRUCTURE: Backley & Nasukawa, 2020). As a result, the structures |AI̲|, |AUI̲|, |UI̲| are realised as palatalized [æː] [øː] [yː] (rather than [aː] [oː] [uː]) (Harris, 1994; Backley, 2011). As a salient property, palatality may yield glide-vowel sequences such as [jaː] as phonetic variants in Nagoya Japanese. This analysis serves as a starting point for exploring other processes that have been accounted for in terms of precedence. The aim is to strengthen the claim that precedence is not a formal structural property anywhere in the language faculty.

## References

Chomsky, N. (2010). Some simple evo devo theses: How true might they be for language? In R. Larson, V. Deprez and H. Yamakido (Eds.), *The evolution of human language: Biolinguistic perspectives* (pp. 45-62). Cambridge: Cambridge University Press.

Backley, P. (2021). Elements and structural head-dependency. In S. Bendjaballah, A. Tifrit, & L. Voeltzel (Eds.), *Perspectives on Element Theory* (pp. 9-31). Boston and Berlin: Mouton de Gruyter.

Backley, P., & K. Nasukawa (2020). Recursion in melodic-prosodic structure. In K. Nasukawa (Ed.), *Morpheme-internal recursion in phonology* (pp. 11-35). Boston and Berlin: Mouton de Gruyter.

Harris, J. (1994). *English sound structure*. Oxford: Blackwell.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569-1579.

Tanaka, S. (2022). Vowel coalescence as head-dependent merge. *Phonological Studies*, 25, 77-88.

# THE COMPLEXITY OF A LANGUAGE IS SHAPED BY THE COMMUNICATIVE NEEDS OF ITS USERS: AN ITERATED HIERARCHICAL BAYESIAN MODEL

VLAD C. NEDELCU*[1] and KENNY SMITH[1]

*Corresponding Author: v.c.nedelcu@sms.ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, UK

Recent experimental and modelling work has found that languages are shaped by the referential context in which they operate (Müller, Winters, and Morin, 2019; Nölle, Staib, Fusaroli, and Tylén, 2018; Lupyan and Dale, 2010). Wray and Grace (2007) argues that even compositionality (the fact that the meaning of an utterance is a function of the meaning of its parts and the way in which these parts are combined) may have only evolved in response to changing social contexts. But how can the referential contexts of individual interactions come to shape the level of compositionality in the language of an entire community?

To explore these questions, we propose an iterated hierarchical Bayesian model that shows how partner-specific linguistic innovations can be generalized as community-wide features via a context-sensitive pathway. We combine insights from the approaches of Winters, Kirby, and Smith (2018) and Hawkins et al. (2021), manipulating reliability of context in pair-based interaction, while also considering abstractions of socio-linguistic knowledge at a higher, population-wide level. More specifically, an agent maintains a separate partner-specific representation of context for each communication partner, but a single overhypothesis about the distribution of types of contexts within the community, and a single community-wide distribution of possible languages. This setup allows an agent to dynamically adapt its language use to the communicative needs of individual partners, but also offers a mechanism for generalization of knowledge across different partners. We model communication between agents as an asymmetric reference game, where each agent is assigned the role of speaker or listener. Each game is split into multiple rounds in which the speaker is provided with a target meaning to convey and the listener is confronted with a context consisting of that target plus a number of distractors. An agent's type of context determines the semantic features in which the referents in any given context can differ for that agent. In a *one-feature-different context-type*, a single feature is sufficient for discriminating among all the referents in a context, and that feature is furthermore consistent across all rounds. In a *mixed context-type*, the meanings in any given context also differ in one feature, but that feature differs on a round-by-round basis. Following the iterated learning framework (Kirby, 2001), agents in a generation observe data produced by speakers in the previous generation, then communicate with one another, with these processes being repeated for a large number of generations.

We first simulate a single-partner variant to evaluate our model's predictions against the experimental results of Winters, Kirby, and Smith (2015). In the *shared context* version, we assume that the speaker has direct access to the listener's context. We compare results to the *unshared context* version, where the speaker has uncertainty both about the context-type of the listener and which specific context the listener faces on any given round. In case of a shared context and one-feature context-type, we find the emergence of *one-feature languages* (i.e., which map all meanings sharing one of the features to the same signal), as these are the simplest communicatively functional in one-feature contexts. In the shared mixed context-type condition, the optimal strategy that guarantees communicative success leads to the emergence of *compositional languages* (i.e., which have consistent mappings for all features that make up the meaning). When context is not shared, speakers must infer the context-type of the listener over the course of interaction. This introduces an additional level of uncertainty, and it becomes a possibility for the speaker to design their utterance for the wrong context-type. Speakers are thus encouraged to produce more autonomous and less context-dependent utterances, generally resulting in the emergence of a higher proportion of compositional systems than in the case of a shared access to context.

We then examine the results for the multi-partner model. Where the speaker's communicative partners are homogeneous in facing one-feature or mixed contexts, the hierarchical model produces a similar pattern of results to the single-partner simulations: we find the emergence of languages that best compromise between ease of learning, adaptation to the homogeneous context-type, and robustness to uncertain inferences. However, we also find that the successive swapping of partners causes more overall uncertainty over context-types compared to the single-partner simulations, resulting in the emergence of more compositional languages. Furthermore, when a speaker interacts with partners with heterogeneous context-types, we find additional effects of this heterogeneity: when encountered partners have one-feature contexts, but differ in the specific feature that needs to be encoded for disambiguation, compositional languages become dominant, as speakers must compromise on using a language that encodes both features.

In sum, we show that the degree of compositionality that evolves in the language of a community is dependent on the communicative needs of its members, but also on the degree of uncertainty over the nature of those needs. Our model also tests the theory proposed by Wray & Grace for the emergence of compositionality, and finds that compositionality can emerge in communities where simpler languages would satisfy the individual needs of its users, as long as the community's needs as a whole are heterogeneous.

## Acknowledgements

This research received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant 681942, held by K. Smith), and from a University of Edinburgh College of Arts, Humanities and Social Sciences Research Award (held by V. Nedelcu).

## References

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2021). From partners to populations: A hierarchical bayesian account of coordination and convention. *arXiv preprint arXiv:2104.05857*.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1).

Müller, T. F., Winters, J., & Morin, O. (2019). The influence of shared visual context on the successful emergence of conventions in a referential communication task. *Cognitive science*, *43*(9), e12783.

Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, *181*, 93–104.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(3), 415–449.

Winters, J., Kirby, S., & Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition*, *176*, 15–30.

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*(3), 543–578.

# INTEGRATION AND SEGREGATION OF AUDITORY STREAMS DURING RHYTHM PERCEPTION

JUN NITTA*[1], SOTARO KONDOH[1], KAZUO OKANOYA[1,2], and RYOSUKE O. TACHIBANA[1],

*Corresponding Author: nitta-jun477@g.ecc.u-tokyo.ac.jp
[1]Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
[2] Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan

Among auditory information, rhythm is one of the important factors in human language (Tachibana et al., 2013), and in vocalization of other species (Fitch, 2013). In the present study, we examined the rhythm perception of human, especially in relation with the effect of timbre.

Research on rhythm perception has mainly focused on temporal structure, including periodicity (Bendixen et al., 2010). However, research on the auditory stream segregation has reported that sounds with different timbres are difficult to perceive as a single unit (Bregman, 1990). This suggests that rhythmic sequences composed of multiple timbres might be more difficult to discriminate their rhythmic patterns than those composed of single timbre.

In the present study, we aimed to reveal the perception of rhythmic pattern including different timbres, and investigated the accuracy of rhythm discrimination by 2 Alternative Forced Choice Task. The center frequency of the band noise stimuli was manipulated to provide conditions with and without frequency difference. In the experiment, participants listened to two rhythmic stimuli, standard stimuli and comparative stimuli, and were asked to choose the bouncing one. Point of subjective equality (PSE) and discrimination accuracy ($\sigma$) were estimated by sigmoid approximation.

The result (Fig. 1) showed that the accuracy of rhythm discrimination decreased in the condition with frequency difference. A paired t-test for $\sigma$ showed significant differences between conditions ($p < 0.001$, $t = 5.36$, $df = 39$). There effects were found even when the standard rhythm pattern was varied from 2:1, 1:2, to 1:1. This suggests that rhythm perception is affected not only by temporal structure but also by sequential grouping. Similar experiments on other species would give an insight into the origin of language.

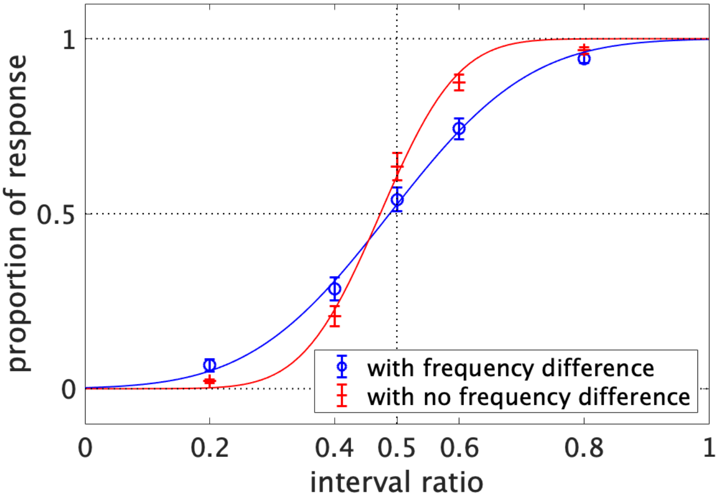


Figure 1. Proportion of response that the comparative stimulus was "bouncing" for each interval ratio. The blue line indicates the condition with frequency difference, and the red line indicates the condition without frequency difference. Error bars are standard errors. Curves are sigmoidal approximations.

## Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP20K21804.

## References

Bendixen, A., Denham, S. L., Gyimesi, K., & Winkler, I. (2010). Regular patterns stabilize auditory streams. The Journal of the Acoustical Society of America, 128(6), 3658-3666.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, 7, 68.

Tachibana, R. O., Sasaki, Y., & Riquimaroux, H. (2013). Relative contributions of spectral and temporal resolutions to the perception of syllables, words, and sentences in noise-vocoded speech. *Acoustical Science and Technology*, *34*(4), 263–270.

# MORE THAN AFFECT: HUMAN FACIAL EXPRESSIONS PROVIDE ICONIC AND PRAGMATIC FUNCTIONS

JONAS NÖLLE*[1], YICHEN WU[1], PABLO ARIAS[2,3], OLIVER G. B. GARROD[1],
PHILIPPE G. SCHYNS[1] and RACHAEL E. JACK[1]

*Corresponding Author: jonas.noelle@glasgow.ac.uk
[1]School of Psychology & Neuroscience, University of Glasgow, UK
[2]Cognitive Science, Lund University, Sweden
[3] Sound Perception and Design, IRCAM, Paris, France

Human facial expressions are a powerful tool for social communication. In addition to the visually salient appearance of the human face (Tomasello et al., 2007; Vick et al., 2007), it has a high capacity to generate many different facial expressions that can each represent different nuanced social messages. Though used daily for a variety of social interactions, facial expressions have mainly been studied as displays of affect in humans (Ekman, 1994) and nonhuman primates (Van Hooff, 1972; Waller & Micheletta, 2013) and less so as a pragmatic tool for communication (cf. Crivelli & Fridlund, 2018). Thus, while multimodal accounts of communication continue to gain attention (e.g., Perniss, 2018; Holler & Levinson, 2019), including in discussions of language origins (Zlatev et al., 2017; Fröhlich et al., 2019), the contribution of facial expressions to multimodal communication remains understudied. One main challenge to this endeavor is the sheer number and complexity of facial expressions (Jack et al., 2018).

To unravel this complexity, we used a powerful data-driven method (Yu et al., 2012) that agnostically generates facial movements and objectively measures their effect on social perception (see Jack & Schyns, 2017 for a review). Here, we examined the semiotic potential of facial expressions combined with speech as part of multimodal signaling, addressing two questions: whether facial expressions can 1) express meanings iconically (e.g., representing size) and/or 2) serve as pragmatic markers, akin to speech prosody, described below.

**Experiment 1.** We tested whether facial expressions can communicate iconic meanings, similar to manual gestures, e.g., pinching to express 'tiny' numbers referred to in speech (Woodin et al., 2020). We presented participants (20 native English speakers, 10 females) with a scenario (e.g., *"There are 108 animals in the field"*) followed by a series of multimodal stimuli—i.e., speakers each displaying a facial expression and commenting on the scenario one of three vague quantifiers *("Of these, several/many/few were cows")*. Participants viewed each stimulus and estimated the number communicated by the speaker. Facial expressions displayed on each trial comprise combinations of eye/eyebrow movements (called Action Units—AUs; Ekman & Friesen, 1978) that do not affect speech, agnostically generated by a generative model of real human facial movements (Yu et al., 2012). We generated different voices using recordings of two native English speakers, normalized and transformed via pitch and spectral envelope shifts to create naturalistic but tightly controlled speaker identities (Arias et al., 2021). We generated synchronized lip movements using a neural net (Cudeiro et al., 2019) and aligned the onset of each facial expression with the spoken quantifier. Finally, we displayed each facial expression on a randomly generated face identity using a generative model of human face shape/complexion (Zhan et al., 2019). Analysis of the statistical relationship between the facial movements presented on each trial and each participant's responses (Ince et al., 2017) revealed that specific facial movements modulate the quantities estimated by the participants.

**Experiment 2.** Using the same data-driven approach, we examined the pragmatic function of facial expressions to mark confidence or doubt, also known as 'Feeling of Knowing,' akin to prosodic cues when answering questions (Jiang & Pell, 2017). Here, the scenario comprised a question (e.g., "*Is she a good leader?"*) followed by a multimodal stimulus (i.e., speakers displaying a facial expression) responding 'yes' or 'no.' Participants rated the perceived confidence of the speaker's answer using a 5-point scale ('very doubtful' to 'very confident'). Analysis using ordinal logistic regression revealed that specific facial movements influence the perception of confidence (AU43 eyes closed, AU1-2 brow raiser) and doubt (AU4 brow lowerer, AU7 lid tightener).

In sum, we used a data-driven psychophysical approach combined a with generative model of the human face to precisely characterize the facial movements that serve pragmatic functions in multimodal communication. Our results suggest that facial movements originally evolved to control sensory input (Susskind et al., 2008) can be exapted to ground more abstract social signals or iconic representations, with direct implications for central theories of multimodal communication and language origins (e.g., Holler & Levinson, 2019).

## Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. [759796]) to REJ; Wellcome Trust [Senior Investigator Award; 107802], Multidisciplinary University Research Initiative/Engineering and Physical Sciences Research Council [172046-01] awarded to PGS; the Swedish Research Council [2014-1371] awarded to PA. We would also like to thank Bodo Winter for discussions about the design, Philip McAleer and Gaby Mahrholz for help with sound recordings, as well as Niamh Mary McLaughlin and Jude Kiltie, who kindly lent us their voices.

## References

Arias, P., Rachman, L., Liuni, M., & Aucouturier, J.-J. (2021). Beyond Correlation: Acoustic Transformation Methods for the Experimental Study of Emotional Voice and Speech. *Emotion Review*, *13*(1), 12–24. https://doi.org/10.1177/1754073920934544

Crivelli, C., & Fridlund, A. J. (2018). Facial Displays Are Tools for Social Influence. *Trends in Cognitive Sciences*, *22*(5), 388–399. https://doi.org/10.1016/j.tics.2018.02.006

Cudeiro, D., Bolkart, T., Laidlaw, C., Ranjan, A., & Black, M. J. (2019). *Capture, Learning, and Synthesis of 3D Speaking Styles*. 10101–10111. https://openaccess.thecvf.com/content_CVPR_2019/html/Cudeiro_Capture_Learning_and_Synthesis_of_3D_Speaking_Styles_CVPR_2019_paper.html

Ekman, P. (1994). All emotions are basic. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 65–58). Oxford University Press.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Investigator's guide*. Consulting Psychologists Press.

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & Schaik, C. P. van. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809–1829. https://doi.org/10.1111/brv.12535

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Ince, R. A. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human Brain Mapping*, *38*(3), 1541–1573. https://doi.org/10.1002/hbm.23471

Jack, R. E., Crivelli, C., & Wheatley, T. (2018). Data-Driven Methods to Diversify Knowledge of Human Psychology. *Trends in Cognitive Sciences*, *22*(1), 1–5. https://doi.org/10.1016/j.tics.2017.10.002

Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, *68*(1), 269–297. https://doi.org/10.1146/annurev-psych-010416-044242

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126. https://doi.org/10.1016/j.specom.2017.01.011

Perniss, P. (2018). Why We Should Study Multimodal Language. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01109

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, *11*(7), 843–850. https://doi.org/10.1038/nn.2138

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, *52*(3), 314–320. https://doi.org/10.1016/j.jhevol.2006.10.001

Van Hooff, J. (1972). A comparative approach to the phylogeny of laughter and smiling. *Non-Verbal Communication*, 209–241.

Vick, S.-J., Waller, B. M., Parr, L. A., Smith Pasqualini, M. C., & Bard, K. A. (2007). A Cross-species Comparison of Facial Morphology and Movement in Humans and Chimpanzees Using the Facial Action Coding System (FACS). *Journal of Nonverbal Behavior*, *31*(1), 1–20. https://doi.org/10.1007/s10919-006-0017-z

Waller, B. M., & Micheletta, J. (2013). Facial Expression in Nonhuman Animals. *Emotion Review*, *5*(1), 54–59. https://doi.org/10.1177/1754073912451503

Woodin, G., Winter, B., Perlman, M., Littlemore, J., & Matlock, T. (2020). 'Tiny numbers' are actually tiny: Evidence from gestures in the TV News Archive. *PLOS ONE*, *15*(11), e0242142. https://doi.org/10.1371/journal.pone.0242142

Yu, H., Garrod, O. G. B., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Computers & Graphics*, *36*(3), 152–162. https://doi.org/10.1016/j.cag.2011.12.002

Zhan, J., Garrod, O. G. B., van Rijsbergen, N., & Schyns, P. G. (2019). Modelling face memory reveals task-generalizable representations. *Nature Human Behaviour*, *3*(8), 817–826. https://doi.org/10.1038/s41562-019-0625-3

Zlatev, J., Wacewicz, S., Żywiczyński, P., & van de Weijer, J. (2017). Multimodal-first or pantomime-first?: Communicating events through pantomime with and without vocalization. *Interaction Studies*, *18*(3), 465–488. https://doi.org/10.1075/is.18.3.08zla

# NEURAL SIGNATURES OF PROSODIC PROCESSING

CHANTAL ODERBOLZ*[1], ELISABETH STARK[2,3], and MARTIN MEYER[1]

*Corresponding Author: chantal.oderbolz@uzh.ch
[1]Department of Comparative Language Science, University of Zurich, Zurich, Switzerland
[2]Zurich Center for Linguistics, University of Zurich, Zurich, Switzerland
[3]Institute of Romance Studies, University of Zurich, Zurich, Switzerland

## 1. Introduction

Generating complex meaning from continuous speech requires the neural mapping of this incoming signal onto syntactic structure. In the evolution of language, prosody might have been integral in facilitating this mapping in the human brain – essentially providing a physical nexus to the abstract (Kreiner & Eviatar, 2014). Support for the possible phylogenetic significance of prosody comes from research on human ontogeny: Infants use a range of prosodic cues to infer different aspects of language in a process called prosodic bootstrapping (Hawthorne & Gerken, 2014). In trying to elucidate this mapping in adults, we refer to cortical oscillations as the biophysical infrastructure that might be suited to enabling the transformation from the physical to the abstract (Ding & Simon, 2014; Giraud & Poeppel, 2012; Meyer, 2018; Murphy, 2020). More precisely, cortical oscillations are hypothesized to synchronize with the speech signal – a process referred to as speech tracking. However, the extent to which this is functionally relevant for continuous speech processing and which prosodic features serve as driving forces is still not fully understood.

## 2. Methods

We therefore sought to better understand the dynamics of speech tracking and the extent to which it is modulated by prosodic features. To this end, we conducted an EEG experiment with a sample of native Swiss German speakers (N = 26). Participants listened to spoken sentences manipulated in terms of their rhythm and intonation. Roughly the first half of a sentence could either follow a predictable regular or an unpredictable irregular rhythm. The second half then either confirmed these predictions or violated them. Additionally, sentences either had a natural or pitch-flattened intonation contour. To test whether these manipulations were perceptually relevant, sentences were embedded in noise. We expected the quality of speech tracking to be influenced by the prosodic features of speech and this relationship between speech signal and brain signal to predict speech perception. To test these hypotheses, we calculated generalized linear mixed models. Specifically, we assessed the phase-locking value (Lachaux et al., 1999) as an indication of speech tracking and percentage of correctly identified syllables in noise as a perceptual marker.

## 3. Results

Results showed that the brain is sensitive even to slight variations in rhythmic (ir)regularity. Cortical oscillations in the theta range (3.5–4.7 Hz) were able to track the speech signal more faithfully when the rhythm was regular compared to irregular. However, the predictability of this rhythm did not seem to influence speech tracking: Neither the violation of a predictable rhythm nor the continuation of an unpredictable rhythm caused a reduction in the PLV. In contrast to rhythm, theta oscillations were less sensitive to intonation as evidenced by the equally high PLV values in either intonation condition.

## 4. Discussion

Continuous speech tracking was differentially influenced by the prosodic features of speech. A regular rhythm — as compared to an irregular rhythm — facilitated the brain’s ability to phase-align with speech in the theta range. These findings are in accordance with studies showing advantages in evoked potentials and perception afforded by rhythmic regularity in auditory stimuli (Abecasis et al., 2005) as well as speech (Bohn et al., 2013; Roncaglia-Denissen et al., 2013). Underlying this facilitation is presumably an evolutionarily rooted connection between the speech motor system, auditory perceptual system, and neural infrastructure (Giraud & Poeppel, 2012; Poeppel & Assaneo, 2020; Strauss & Schwartz, 2017). As such, rhythm may have evolved to take on the role of an endogenous tool, structuring speech input and optimizing sensory processing. Intonation, on the other hand, was not related to differences in speech tracking. Given this non-existent effect, it appears that these two prosodic features and their respective functions are dissociable in and dissociated by the brain and thus might have also had different evolutionary trajectories. Along these lines, intonation may come into play for more strictly high-level linguistic operations such as syntactic parsing or disambiguation (Marslen-Wilson et al., 1992; Selkirk, 2011).

## Acknowledgements

This work was supported as a part of NCCR Evolving Language, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number #51NF40_180888).

## References

Abecasis, D., Brochard, R., Granot, R., & Drake, C. (2005). Differential brain response to metrical accents in isochronous auditory sequences. *Music Perception, 22*(3), 549-562. doi:https://doi.org/10.1525/mp.2005.22.3.549

Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir)regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia, 51*(4), 760-771. doi:https://doi.org/10.1016/j.neuropsychologia.2013.01.006

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience, 8*(311). doi:https://doi.org/10.3389/fnhum.2014.00311

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*, 511. doi:https://doi.org/10.1038/nn.3063

Hawthorne, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2), 420–428. https://doi.org/10.1016/j.cognition.2014.07.013

Kreiner, H., & Eviatar, Z. (2014). The missing link in the embodiment of syntax: Prosody. *Brain and Language*, *137*, 91–102. https://doi.org/10.1016/j.bandl.2014.08.004

Lachaux, J.-P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping, 8*(4), 194-208. doi:10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology Section A, 45*(1), 73-87. doi:https://doi.org/10.1080/14640749208401316

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience, 48*(7), 2609-2621. doi:https://doi.org/10.1111/ejn.13748

Murphy, E. (2021). *The oscillatory nature of language*. Cambridge University Press.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21*(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLOS ONE, 8*(2), e56000. doi:https://doi.org/10.1371/journal.pone.0056000

Selkirk, E. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (2nd ed., pp. 435-484). West Sussex, UK: Blackwell Publishing Ltd.

Strauß, A., & Schwartz, J.-L. (2017). The syllable in the light of motor skills and neural oscillations. *Language, Cognition and Neuroscience, 32*(5), 562-569. doi:https://doi.org/10.1080/23273798.2016.1253852

# NEUROFEEDBACK TRAINING AS A METHOD FOR EVALUATION OF BRAIN SYSTEMS FOR LANGUAGE

JIRO OKUDA*[1], HIROKI KATAOKA[1], MASAYUKI FUJIWARA[2], and TAKASHI HASHIMOTO[2]

*Corresponding Author: jokuda@cc.kyoto-su.ac.jp
[1]Division of Frontier Informatics, Kyoto Sangyo University, Kyoto, Japan
[2]School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan

## 1. Introduction

Since language is a complex faculty concerning multiple domains, it is thought that the brain system for language has evolved in an integrative manner such that numbers of componential systems make their own contributions to the whole system. Recently, on the other hand, neuroscience studies have begun to show that a neurofeedback training of repeated self-induction of a specific brain activity pattern could alter behavioral ability corresponding to the trained brain activity (Shibata *et al.*, 2011). Given this causal effect, we wondered that the neurofeedback training might serve as a new method to test causal relevance of hypothetical neuro-linguistic models. That is, if behavioral performance on a certain linguistic ability changed according to the self-induction training of activity of a specific brain system, we could argue that such brain system might have effective contribution to that ability in an adaptive manner.

Based on this idea, we specifically tried to examine organization and possible adaptive capacity of brain systems contributing to an ability to express intentions by hierarchical structures of written symbol strings. Although both the hierarchical structure (Everaert *et al.*, 2015) and the intention-sharing (Tomasello, 2003) are believed to be essential for human language, relation between the two remains unknown. Therefore, we first proposed a neural model concerning intention-sharing by the hierarchical structures of symbol strings through an electroencephalography (EEG) experiment, and then tested the effect of the EEG neurofeedback training with respect to the model.

## 2. EEG experiment and neurofeedback training

We adopted experimental semiotics approach (Galantucci, 2009) that could simulate emergence and evolution of a novel language system through trial and error across participants. In each trial of our experiment, we assigned a pair of participants either a sender or a receiver of a string of three symbols involving the hierarchical structures. The sender composed a symbol string to indicate a target object in the experimental screen, and was additionally required to imply an instruction whether the receiver should choose the target or not. As a result, most participants expressed the meaning of the choose/not-choose instruction within specific hierarchical structures of the string. EEG analyses showed significant suppression of 11 Hz EEG power at the left frontal, bilateral parietal, and occipital electrodes during a composition period of the symbol string. Thus we proposed a neuro-linguistic model that desynchronization of alpha-band neural activity in multiple brain systems (Murphy, 2020) would contribute to realization of the intention expression through the hierarchical structures.

We conducted 4-days neurofeedback training (1.5h per day) to independent participants. We visually presented feedback signals that represented magnitude of 11 Hz power suppression at the frontal, parietal, and occipital electrodes, and asked the participants to try to increase the feedback signals as high as possible. Although effects of the neurofeedback training on performance change between pre- and post-training behavioral tests varied across participants, we observed the representative case in which richness of intention-reading in a picture description task (Cummings, 2019) dramatically improved after successful training of the EEG suppression at the parietal and occipital electrodes. This case also showed less improvement of utilization of the hierarchical structures to express intentions in the string composition task, in parallel with less achievement of the left frontal power suppression during the EEG training.

## 3. Discussion

Differential training effects across pragmatic understanding of intentions and utilization of the hierarchical structures may suggest partially independent evolutionary organization of brain systems for these aspects of language. We speculate that incorporating connectivity information into the neurofeedback method might help to further elucidate integrative organization of these systems.

## Acknowledgements

Supported by KAKENHI JP (20H05019, 20H04256, 18H05087, 17H06383).

## References

Cummings, L. (2019). Describing the Cookie Theft picture: Sources of breakdown in Alzheimer's dementia. *Pragmatics and Society, 10(2),* 153-176.

Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C., & Bolhuis, J. J. (2015). Structures, not strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Sciences, 19(12),* 729-743.

Galantucci, B (2009). Experimental semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science, 1(2),* 393-410.

Murphy, E. (2020). *The oscillatory nature of language.* Cambridge: Cambridge University Press.

Shibata, K., Watanabe, T., Sasaki, Y., & Kawato, M. (2011). Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. *Science, 334(6061),* 1413–1415.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

# SIMILARITIES AND DIFFERENCES IN A CROSS-LINGUISTIC SAMPLE OF SONG AND SPEECH RECORDINGS

YUTO OZAKI[1], JIEI KUROYANAGI[2], JOHN MCBRIDE[3], POLINA PROUTSKOVA[4], ADAM TIERNEY[5], PETER PFORDRESHER[6], EMMANOUIL BENETOS[4], FANG LIU[7], PATRICK E. SAVAGE[2*]

[*]Corresponding Author: psavage@sfc.keio.ac.jp
[1]Graduate School of Media and Governance, Keio Univ., JP
[2]Faculty of Environment and Information Studies, Keio Univ., JP
[3]Center for Soft and Living Matter, Institute for Basic Science, KR
[4]School of EECS, Queen Mary Univ. of London, UK
[5]Department of Psychological Sciences, Birkbeck, Univ. of London, UK
[6]Department of Psychology, Univ. at Buffalo, NY, US
[7]School of Psychology and Clinical Language Sciences, Univ. of Reading, UK

1. Introduction: Music and language are prominent forms of acoustic communication across human societies. The relationship between these two modes of human sound communication has been investigated from the perspective of music and language evolution (e.g. Fitch, 2006; Patel, 2007). Song and speech are usually considered as distinct categories, but the boundary is not always clear (Brown, 2017; Cummins, 2020; Deutsch et al., 2011; Engelhardt & Bretèque, 2017; Feld & Fox, 1994). Many studies have documented acoustic differences between song and speech in specific languages, but few studies have identified consistent cross-cultural similarities or differences across a diverse set of languages (see the references cited in the Result section). Therefore, we aim to conduct a series of comparative analyses with diverse cross-cultural samples to explore more general acoustic similarities and differences. In particular, we analyzed the following acoustic features: fundamental frequency (F0) of voice, inter-onset interval (IOI) of vowel onsets, ratio of F0 between adjacent IOI segments (interval). Onset annotations were performed manually and intervals were calculated by taking the outer product of an F0 vector at an IOI segment and

the reciprocal of the F0 vector at the previous IOI segment which resulted in the distribution of pitch ratio of adjacent IOI segments[1].

2. Dataset: We analyzed 23 pairs of song and speech recordings (i.e. 46 audio files). Each pair was recorded by the same person. 36 recordings were sampled from Hilton et al. (in press) that include English, Mandarin and Spanish. The remaining 10 recordings were newly collected ones in Japanese, English, Mandarin and Korean. The latter data was created by singing a song first and then reciting the text of the sung lyrics.

3. Results: We observed that song has higher F0 and longer IOI than speech, as reported previously (Hansen et al., 2020; Merrill & Larrouy-Maestri, 2018; Sharma et al., 2021, but see also Ding et al., 2017 who used amplitude modulation). We also observed that song has a sharper concentration at an IOI ratio (Roeske et al., 2020) of 0.5 than speech though both data has a median around 0.5. Our nPVI analysis both failed to support Patel & Daniele (2003)'s song-speech relationship hypothesis and failed to sort languages into traditional syllable-/stress-/mora-timed categories. Speech and song showed similar ranges of melodic intervals (within ±700 cents). However, the interval distribution of song has distinct peaks at around ±200 cents in addition to 0 cent, while speech had only one peak at 0 cents. In addition, we measured the variability of F0 by entropy (like Ozaki et al., 2022), and English and Spanish showed that singing has lower entropy than speech indicating greater pitch stability of singing (Merrill & Larrouy-Maestri, 2018; Raposo de Medeiros et al., 2021; Sharma et al., 2021; Thompson, 2014), while Mandarin showed the opposite pattern[1].

4. Conclusion: In summary, we observed the above potential systematic differences and similarities between song and speech. Why and how these differences have emerged and how evolutionary theories of language and musicality can account for these are the key questions to be addressed in future research (Darwin, 1871; Brown, 2000; Tierney et al., 2011; Savage et al., 2021; Mehr et al., 2021). For example, slower and more regular vocal communication may facilitate synchronization (social bonding hypothesis, Savage et al., 2021), while similar melodic interval range and 1:1 IOI duration focus may be due to shared constraints on the vocalization mechanism (motor constraint hypothesis, Tierney et al., 2011). Future steps to explore more comprehensive and robust relationships between two universal human acoustic communication forms include increasing sample size and language diversity, timbre/formant analysis, and measuring inter-rater reliability for the syllable/note annotations.

---

[1] Further details and figures can be found in the supplementary materials. https://drive.google.com/file/d/1MDnk9miUzRMupb8pMtqlpdcuQ-I7w-CY/view?usp=sharing

## Acknowledgements

We thank Sangbuem Leonard Choo for providing Korean language data. This work was supported by Grant-in-Aid no. 19KK0064 from the Japan Society for the Promotion of Science and by startup grants from Keio University (Keio Global Research Institute, Keio Research Institute at SFC, and Keio Gijuku Academic Development Fund) to P.E.S, and by Grant Number JPMJSP2123 of Support for Pioneering Research Initiated by the Next Generation from the Japan Science and Technology Agency to Y.O.

## References

Brown, S. (2000). The “Musilanguage” model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). MIT Press.

Brown, S. (2017). A Joint Prosodic Origin of Language and Music. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.01894

Cummins, F. (2020). The Territory Between Speech and SongA Joint Speech Perspective. *Music Perception*, *37*(4), 347–358. https://doi.org/10.1525/mp.2020.37.4.347

Darwin, C. (1871). *The descent of man*. Watts & Co.

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. The Journal of the Acoustical Society of America, 129(4), 2245–2252. https://doi.org/10.1121/1.3562174

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, *81*, 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Engelhardt, J., & Bretèque, E. A. de la. (2017). Guest Editors’ Preface: Speech, Song, and In-Between. *Yearbook for Traditional Music*, *49*, xv–xix. https://doi.org/10.5921/yeartradmusi.49.2017.00xv

Feld, S., & Fox, A. A. (1994). Music and Language. *Annual Review of Anthropology*, *23*, 25–53.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173–215. https://doi.org/10.1016/j.cognition.2005.11.009

Hansen, J. H. L., Bokshi, M., & Khorram, S. (2020). Speech variability: A cross-language study on acoustic variations of speaking versus untrained singing. *The Journal of the Acoustical Society of America*, *148*(2), 829. https://doi.org/10.1121/10.0001526

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., Schachner, A., … Mehr, S. A. (in press). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*. https://doi.org/10.1101/2020.04.09.032995

Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *Behavioral and Brain Sciences*, *44*. https://doi.org/10.1017/S0140525X20000345

Merrill, J., & Larrouy-Maestri, P. (2017). Vocal Features of Song and Speech: Insights from Schoenberg's Pierrot Lunaire. *Frontiers in Psychology*, *8*, 1108. https://doi.org/10.3389/fpsyg.2017.01108

Ozaki, Y., Sato, S., Mcbride, J., Pfordresher, P. Q., Tierney, A. T., Six, J., Fujii, S., & Savage, P. E. (2022). Automatic acoustic analyses quantify pitch discreteness within and between human music, speech, and birdsong. *Proceedings of the 10th International Workshop on Folk Music Analysis*. The 10th International Workshop on Folk Music Analysis, Sheffield, United Kingdom.

Patel, A. D. (2007). *Music, Language, and the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195123753.001.0001

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. Cognition, 87(1), B35–B45. https://doi.org/10.1016/S0010-0277(02)00187-7

Raposo de Medeiros, B., Cabral, J. P., Meireles, A. R., & Baceti, A. A. (2021). A comparative study of fundamental frequency stability between speech and singing. *Speech Communication*, *128*, 15–23. https://doi.org/10.1016/j.specom.2021.02.003

Roeske, T. C., Tchernichovski, O., Poeppel, D., & Jacoby, N. (2020). Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology*, *30*(18), 3544-3555.e6. https://doi.org/10.1016/j.cub.2020.06.072

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*. https://doi.org/10.1017/S0140525X20000333

Sharma, B., Gao, X., Vijayan, K., Tian, X., & Li, H. (2021). NHSS: A speech and singing parallel database. *Speech Communication*, *133*, 9–22. https://doi.org/10.1016/j.specom.2021.07.002

Thompson, B. (2014). Discrimination between singing and speech in real-world audio. *2014 IEEE Spoken Language Technology Workshop (SLT)*, 407–412. https://doi.org/10.1109/SLT.2014.7078609

Tierney, A. T., Russo, F. A., & Patel, A. D. (2011). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences*, *108*(37), 15510–15515. https://doi.org/10.1073/pnas.1103882108

# THE ADAPTIVE SIGNIFICANCE OF HUMAN SCLERAL BRIGHTNESS. AN EXPERIMENTAL STUDY.

JUAN PEREA GARCÍA[1], DARIUSZ DANEL[2,], ZDZISŁAW LEWANDOWSKI[3],
SLAWOMIR WACEWICZ*[4]

*Corresponding Author: wacewicz@umk.pl
[1] National University of Singapore, Singapore
[2] Hirszfeld Institute of Immunology and Experimental Therapy, Wrocław, Poland
[3] University School of Physical Education in Wrocław, Poland
[4] Center for Language Evolution Studies, Nicolaus Copernicus University,
Toruń, Poland

## 1. Objective

Several of the most influential language evolution scholars (e.g. Tomasello et al. 2007, Knight and Lewis 2017, cf. also e.g. Hare 2017) have opined that the emergence of the homogeneously pale sclera characteristic of modern humans was a prerequisite for, and turning point in, the evolution of language. However, the understanding of the exact role and origin of this trait has remained highly speculative, with initial investigations suggesting that it evolved to facilitate eye-gaze following by conspecifics. It is unlikely that our species went from the typically pigmented sclera of most other primate species to its current state in the absence of proximate functions that made partial depigmentation adaptive. The literature (see esp. Danel et al. 2020) enumerates several non-mutually exclusive proximate functions that could have mediated the transition from pigmented to depigmented sclerae. These functions include the signaling of:

- attractiveness and health (e.g. through symmetrical depigmentation that affords the perception of changes in vasculature or pigmentation),
- reliability and trustworthiness (e.g. by facilitating the perception of one’s gaze and emotions and, thus, intentions),
- reduced emotional reactivity (possibly sharing proximal mechanisms with self-domestication).The aim of our study was to investigate whether partial scleral depigmentation could have been adaptive by signaling reduced aggressiveness, attractiveness, trustworthiness, health, and age.

## 2. Stimulus and procedure

STIMULUS. We morphed and photo-edited images of the faces of a range of reconstructed hominin species, to create realistic facial images of 20 individually distinct hominins with relatively diverse facial morphologies. We adjusted the brightness of the scleral areas of each individual to achieve a perceptually (i) “humanlike” bright sclera and (ii) “generalised apelike” dark sclera version of each individual. PROCEDURE. 250 demographically diverse participants, recruited via Prolific, rated 10 bright-sclera and 10 dark-sclera hominins randomly presented with online software, by moving sliders between the left (0) and right (100) extremes for each trait. ANALYSIS. We fit separate generalized mixed models (GLMM) for each of our dependent variables (except *Age*, see below) with our manipulation (dark/bright sclerae) as an independent variable. For age, we fit a linear regression model (LM). We included stimuli pairs and participants as random effects in all our models. We predicted that bright-sclera hominin faces would be rated as less aggressive, more attractive, more trustworthy, healthier, and younger than the same faces with a dark sclera. Our study was pre-registered at https://aspredicted.org/TM4_PSK

## 3. Results

All models found statistically significant differences between pairs of stimuli in the predicted direction (Table 1).

Table 1. Summary of the LM (Age) and GLMMs (the remaining dependent variables) for “dark-sclera” vs “bright-sclera” pairs of hominin faces.

| | *Intercept* | *Dark-Bright difference* | *t/z value* | *p* |
|---|---|---|---|---|
| Age (LM) | 63.13 | 4.03 | 5.30 | <0.01 |
| Aggression | -0.44 | 0.28 | 38.369 | <0.01 |
| Attractiveness | -0.19 | -0.37 | -50.76 | <0.01 |
| Health | 0.51 | -0.39 | 53.17 | <0.01 |
| Trustworthiness | 0.24 | -0.31 | -43.25 | <0.01 |

## 4. Conclusions

Our results show that partial scleral depigmentation in hominins may have been adaptive for signaling functions that possibly predate the deictic function of eye-gaze. Further studies are required to evaluate the role of the familiarity (i.e. human-likeness) of the stimulus in the effects detected in our study.

## References

Danel, D. P., Wacewicz, S., Kleisner, K., Lewandowski, Z., Kret, M. E., Żywiczyński, P., & Perea-Garcia, J. O. (2020). Sex differences in ocular morphology in Caucasian people: a dubious role of sexual selection in the evolution of sexual dimorphism of the human eye. Behavioral Ecology and Sociobiology, 74(10), 1-10.

Hare, B. (2017). Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality. *Annual Review of Psychology* 68:1, 155-186.

Knight, C., & Lewis, J. (2017). Wild voices: Mimicry, reversal, metaphor, and the emergence of language. *Current Anthropology*, 58(4), 435-453.

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *Journal of human evolution*, 52(3), 314-320.

# BODILY ROTATION IN THE GESTURING, PLAY, AND DISPLAY OF GREAT APES

MARCUS PERLMAN

m.perlman@bham.ac.uk
Department of English Language and Linguistics, University of Birmingham, Birmingham, UK

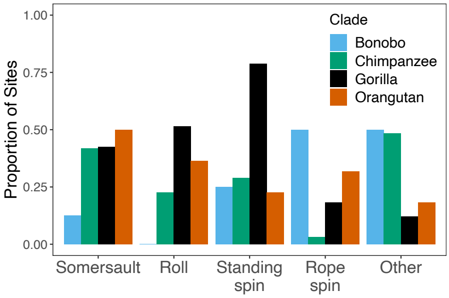
Study of the gesturing and displays of great apes is, arguably, fundamental to understanding the evolution of human language, as well as other forms of human expression like dance. This paper examines the case of bodily rotation in ape gesturing and display, such as in various forms of spinning, rolling, flipping, etc. Such rotational movements are a recurring element in the intentional communication of great apes, for example, noted by Byrne et al. (2017) in six distinct gesture types and variants: *pirouette*, *pirouette with object, ice skating, side roulade, somersault,* and *rope spinning*. Byrne et al. hypothesized that these gestures, and dozens of others, are part of the innate repertoires of ape species, evolved through phylogenetic ritualization. By this idea, although such gestures are used intentionally and flexibly in context, they lack the open-endedness in form that is characteristic of human language. This hypothesis also implies that the human propensity for rotating their bodies—common in creative activities like dance and play—is not related to similar looking behaviors in apes, as it is doubtful our own diverse behaviors stem from a shared inventory of innately specified gestures. Alternatively, humans and great apes might spin, roll, and flip for homologous reasons, perhaps for the stimulating, dizzying effects of these behaviors, which drive their widespread occurrence.

Here, the rotational behavior of great apes was analyzed in a large sample of online videos collected from YouTube. Videos were identified by search phrases that combined terms for primate clades (e.g., chimpanzee, gorilla, ape) with different rotational movements (e.g., spinning, rolling, flipping). Data include 276 videos of great apes performing 793 rotational behaviors comprising 361 single rotations and 442 sequences ranging from 2 to 40 rotations. Video captions available for about half of the videos confirm footage from many different sites. Table 1 shows the number of observations by clade, including the number of behaviors, videos, and caption-confirmed sites. Sites included zoos and sanctuaries, animals handled as pets and trained acts, and animals observed in the wild.

| *Clade* | *Behaviors* | *Videos* | *Zoo/Handled/Wild* |
|---|---|---|---|
| Bonobo | 60 | 18 | 7 / 1 / 0 |
| Chimp | 195 | 49 | 17 / 12 / 2 |
| Gorilla | 332 | 117 | 30 / 0 / 3 |
| Orang | 206 | 92 | 19 / 3 / 0 |

Table 1. Number of observations by clade.

Analysis shows that individuals across species, ages, and sexes performed a range of rotational behaviors, including those previously documented (Byrne et al., 2017), and others like *backflips* and *leaping 360s*. Behaviors were most often performed in contexts of play and display. Figure 1 shows the frequency of major behavior types by clade, measured as the proportion of confirmed sites at which the behavior was observed. For example, gorillas are notable in their frequent performance of *standing spins*—spinning around the vertical bodily axis while standing upright on varying numbers of hands and feet—in about three-quarters of sites. Not visible in this plot is that, although the behaviors are classifiable as types based on certain formal characteristics, they often show great variation and creativity, produced with distinctive movements tuned to particular features of the environment (e.g., ropes, swings, hills, water, floor types, various objects).



different rotational behaviors were observed.

These preliminary results are interpreted as evidence that apes perform rotational behaviors for similar reasons to humans, a primary one appearing to be that they are fun as a result of being perceptually stimulating and dizzying. Similar-looking inventories of rotational behaviors across ape species might result from this shared inclination channeled through the affordances common to our ape bodies. Importantly, the data show that the forms of rotational behaviors are flexible and subject to creativity, suggesting that they are open-ended, perhaps comparable to human language and gesture.

## References

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Animal Cognition*, *20*, 755-769.

# THETA RHYTHM IS WIDESPREAD IN VOCAL PRODUCTION ACROSS THE ANIMAL REALM.

THÉOPHANE PIETTE[*1], CHUNDRA CATHCART[2], CHIARA BABRIERI[3], DIDIER GRANDJEAN[4], ÉLOÏSE DÉAUX[1], ANNE-LISE GIRAUD[1]

[*]Corresponding Author: theophane.piette@unige.ch

[1]Department of Basic Neuroscience, University of Geneva, Geneva, Switzerland

[2]Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland

[3] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

[4]Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

## 1. Introduction

Speech production rate, conserved between 2 to 8 syllables per second, plays a key role in comprehension. Recent studies that highlighted the presence of the same rhythm in vocal productions and mouth movements of non-human primates, suggest that it may be inherited from mastication (Risueno-Segovia et al, 2020). However, other factors could constrain vocal rhythm. Notably, syllabic rate also matches the frequency of theta brain oscillations (2 -7 Hz). While oscillations are rhythmic electrical activity spontaneously generated by neural tissues, they can also be stimulus-driven and research has demonstrated that theta oscillations entrain to speech during listening (Giraud et al, 2014). As brain oscillations frequencies are conserved in animals, syllabic rate may instead be constrained by the recipient's neuronal activity (Buzaski et al, 2013). To determine what factor(s) may have contributed to the emergence of the theta vocal rhythm, we collected vocalizations for 90 species of animals. We used phylogenetic regression to assess whether weight, mastication or environmental constraints explain the variation in animal vocal rhythms while controlling for phylogenetic relationships.

## 2. Methods

### *2.1 Vocal, physical and environmental data*

We gathered, when they were available, recordings for one species per infra-order of tetrapods as well as some fishes and insects (N=92, mammal = 24, bird = 58, amphibian = 4, reptile =1, insect = 4 and fish= 1). To obtain vocal production rhythm, we then extracted sequences, consisting of one subject vocalizing for at least 2s, with less than 1s of silence between two sounds. For each sequence, the envelope, was computed, and analyzed using a wavelet transform to extract the peak rhythmic changes of amplitude (Tilsen et al 2014). Bodyweight data were collected (N=92), habitat type was defined as either closed, semi-closed, open, wetland or water, and mastication status as yes or no. We obtained data on the context of production, but these were only available for 35% of the recordings. Therefore, we assumed that such effect would relate to flexible rhythm adjustment within an already physically constrained rhythmic range (Martin et al, 2022).

### *2.2 Statistical analyses*

Given the extent of the current dataset, we are still in the process of building the corresponding phylogenetic tree. Thus using log-transformed weight, mastication, environments, and their interactions as fixed effects, and random slopes of log weight by groups and random intervals for orders, we built a linear mixed model investigating the variation of the median log-transformed rhythm frequency.

## 3. Results

We can observe in our preliminary results that most of our species (74.4%) have a median frequency located between 2Hz to 7Hz. When testing our linear mixed model, we noticed that our random effects did not explain the variance of rhythm in our species. We also found that none of our predictors significantly influenced vocal rhythm (Most impacting factor: Environment, F=1.0570, p.value=0.3748).

## 4. Conclusion

Theta production rhythm seems widespread across animal vocal productions, independently of their physical characteristics, mastication status or environment. As it is present in most species, even the ones that do not masticate, it appears unlikely that this rhythm is constrained by a limit of the natural movement of the articulators. Furthermore, the lack of effect of weight and environmental constraints tends to point towards a more common and endogenous constraint for vocalization rhythm. As brain oscillations and neurons characteristics, in general, are well conserved across species, it could be that the rhythm of the vocalizations has adapted to them as a form of ‘neuronal exploitation’, to ensure the efficient reception by the listener. As the range of oscillations in animals’ brains is not documented for most species, further work on phylogeny and the direct link between vocalizations and brain oscillations in animals will be needed to strengthen this hypothesis or lead the way for alternative explanations.

## Acknowledgements

The NCCR Evolving Language, Swiss National Science Foundation Agreement Nr. #51NF40_180888, funded this work. The authors are grateful to the CSIRO Australian National Wildlife Collection, (https://ror.org/059mabc80), Fonoteca Zoologica (https://www.fonozoo.com/), the Museum für Naturkunde of Berlin (https://www.museumfuernaturkunde.berlin) and the Xeno-Canto Foundation for Nature Sounds (https://xeno-canto.org) for their assistance in undertaking this research. The authors thank all the researchers from the NCCR Evolving Language as well as Adrien Meguerditchian, Pascal Belin and Steffen R.Hage for their insights on the project. The authors would like to thank Aaron Bauer, Andrew Spencer, Alex Kwet, Alex Rohtla, Camila Ferrara, Daniel Blumstein, Élodie Briefer, Émilie Genty, Frank Lambert, Isabelle Charrier, Gerald Carter, Hans Schneider, Marc D. Hauser, Nikola Falk, Peter Boesman and Robert Seyfarth for providing us with recordings for this study. The authors also thank two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

## References

Buzsáki, G., Logothetis, N., & Singer, W. (2013). Scaling brain size, keeping timing : evolutionary preservation of brain rhythms. Neuron, 80(3), 751-764. https://doi.org/10.1016/j.neuron.2013.10.002

Fujii, S., & Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: the sep hypothesis. Frontiers in Human Neuroscience, 8. https://doi.org/10.3389/fnhum.2014.00777

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nature Neuroscience, 15(4), 511-517. https://doi.org/10.1038/nn.3063

Hoyo, J. del, Elliott, A., Sargatal, J., & Cabot, J. (Éds.). (1992). Handbook of the birds of the world. Lynx Edicions.

Martin, M., Gridley, T., Elwen, S. H., & Charrier, I. (2022). Feel the beat: Cape fur seal males encode their arousal state in their bark rate. The Science of Nature, 109(1), 5. https://doi.org/10.1007/s00114-021-01778-2

Noujaim, S. F., Lucca, E., Muñoz, V., Persaud, D., Berenfeld, O., Meijler, F. L., & Jalife, J. (2004). From mouse to whale : a universal scaling relation for the pr interval of the electrocardiogram of mammals. Circulation, 110(18), 2802-2808. https://doi.org/10.1161/01.CIR.0000146785.15995.67

Ravosa, M. J., Ross, C. F., Williams, S. H., & Costley, D. B. (2010). Allometry of masticatory loading parameters in mammals. The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, 293(4), 557-571. https://doi.org/10.1002/ar.21133

Risueno-Segovia, C., & Hage, S. R. (2020). Theta synchronization of phonatory and articulatory systems in marmoset monkey vocal production. Current Biology, 30(21), 4276-4283.e3. https://doi.org/10.1016/j.cub.2020.08.019

Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. The Journal of the Acoustical Society of America, 124(2), EL34-EL39. https://doi.org/10.1121/1.2947626

Wilson, D. E., Mittermeier, R. A., & Cavallini, P. (Éds.). (2009). Handbook of the mammals of the world. Lynx Edicions: Conservation International: IUCN.

Worthington, J., Young, I. S., & Altringham, J. D. (s. d.). Short communication the relationship between body mass and ventilation rate in mammals. 4.

# HOW SENSES ARE GROUPED WITHIN WORD FORMS OVER TIME: AN ANALYSIS OF DIACHRONIC COLEXIFICATION PATTERNS

Lucía Pitarch*[1] and Thomas Brochhagen[2]

*lpitarch@unizar.es
[1]Universidad de Zaragoza
[2]Universitat Pompeu Fabra

Languages commonly express multiple meanings with single forms (Piantadosi, Tily, & Gibson, 2012). However, while ambiguity is pervasive, the reasons why some meanings are more likely to be conflated than others –i.e., why they *colexify* more often (François, 2008)– are still under investigation (Xu, Duong, Malt, Jiang, & Srinivasan, 2020; Brochhagen & Boleda, 2022).

Originally the term *colexification* referred to several senses sharing the same word form (e.g. the senses of *right* as 'direction' and 'being correct' colexify). Later studies broadened this term to apply to concepts instead of senses (Xu et al., 2020, e.g.), relying on CLICS3 (Rzymski et al., 2020), the largest database on cross-linguistic colexifications available to date. CLICS3 provides English glosses for meanings lexicalized in source languages (e.g., *sentir* in Catalan expresses both FEEL and LISTEN). When different glosses (e.g., LISTEN and FEEL) share the same word form in the original language, as *sentir* in Catalan, those concepts colexify. This approach enabled many quantitative studies by being more easily and automatically computable. However, relying on English as meta-language leads to a decrease in the granularity of the original term (François, 2008, cf.). For the present study, we linked the data from CLICS3 for romance languages to BabelNet (Navigli & Ponzetto, 2012), one of the largest multilingual sources for structured semantic data. In BabelNet words are split into senses which have a universal identifier shared by all languages. We searched pairs that colexified in CLICS3 in a certain language in BabelNet, and retrieved the information related to the BabelNet senses shared by both colexifying wordforms in CLICS3. By doing so we describe colexifications not only at the concept level, nor through English as a mediator, but rather at the sense level with universal sense abstractions. Beyond enriching the available data, this study offers a novel, quantitative and diachronic, perspective on colexification. It asks what form- and concept-based information is predictive of the maintenance or loss of colexifications over time. For instance, while the concepts FEEL and LISTEN did not colexify in Latin (with the different word forms *sentire* and *audire* respectively), in Catalan they do, and the same word form *sentir* is used to refer to both concepts. Yet, in other languages such as Spanish, Latin's semantic organization is maintained.

We focus on the following intralinguistic features: word length, form confusability,[1] part of speech, number of lexifications per sense (in how many other words of a language it appears) and semantic relations (e.g., part/whole and subsumption). The data was processed by grouping senses belonging to the same word forms into pairs, and then fit using logistic regression models with the above features as predictors and, as response, whether a given colexification from Latin was maintained diachronically in at least one of its daughters. That is, whether sense pairs expressed in Latin with the same word are also still colexified in other romance languages. We fit two different models. In the smallest, more imbalanced, model, encompassing all 782 colexifications annotated with semantic relations, only semantic relation information and form confusability were estimated to impact prediction. In fact, the information that two senses are semantically related (no matter the type of relation) was so informative that all colexifications were predicted as maintained, incurring a 10% error rate. The larger model covers all 406777 pairs, but does not include relational information. The most informative feature for diachronic colexification maintenance in this model was form confusability, followed by the number of lexifications of the synset. The rarer a synset's lexification is, the more colexifications it appears in will be maintained. In both models the more different a word is, on average, from other words in its lexicon –i.e., the less mistakable it is– the more the colexifications it hosts are maintained. Both part of speech and word length were the least informative predictors for diachronic colexification maintanance. Word length's lesser impact compared to other features may be due to its overlap with form confusabillity. As for part of speech, Latin's structure may play a role, as its syntactic functions are mainly expressed by case rather than by functional words (e.g., prepositions). Further study of other language families with different peculiarities should shed further light on these matters. The accuracy obtained with both cross-validated models lies between 80% and 90%.

The enrichment of BabelNet data is ongoing (e.g., (Declerck & Bajčetić, 2021) for phonology or (Nayak, Majumder, Goyal, & Poria, 2021) for relation extraction). In the future we aim to improve our diachronic research with this new information to gain a deeper insight on linguistic change.

## Acknowledgements

This research has been partially supported by DGA/FEDER.

---

[1] Mean Levenshtein distance between a form and the lexicon it draws from. The latter was estimated by scraping unique words from 1250 random wikipedia pages for each language. Initially we focus on orthography but we are currently working on enriching the data with phonological resources.

## References

Brochhagen, T., & Boleda, G. (2022). *When do languages use the same word for different meanings? The Goldilocks principle in colexification.* Cognition.

Declerck, T., & Bajčetić, L. (2021). Towards the addition of pronunciation information to lexical semantic resources. In *Proceedings of the 11th global wordnet conference* (pp. 284–291). University of South Africa (UNISA): Global Wordnet Association.

François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In *Studies in language companion series* (pp. 163–215). John Benjamins Publishing Company.

Navigli, & Ponzetto. (2012). Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, *193*, 217–250.

Nayak, T., Majumder, N., Goyal, P., & Poria, S. (2021). Deep neural approaches to relation triplets extraction: A comprehensive survey. *Cognitive Computation*, *13*(5), 1215–1232.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291.

Rzymski, C., Tresoldi, T., Greenhill, S. J., Wu, M.-S., Schweikhard, N. E., Koptjevskaja-Tamm, M., Gast, V., Bodt, T. A., Hantgan, A., Kaiping, G. A., et al.. (2020). The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data*, *7*(1), 1–12.

Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. *Cognition*, *201*.

# RE-EVALUATING HOCKETT'S DESIGN FEATURES FROM A COGNITIVE AND NEUROSCIENCE PERSPECTIVE: THE CASE OF DISPLACEMENT

MICHAEL PLEYER*[1,2] and ELIZABETH QING ZHANG[1,2]

*Corresponding Author: pleyer@umk.pl

[1] Center for Language Evolution Studies, Nicolaus Copernicus University in Toruń, Toruń, Poland

[2] University Center of Excellence IMSErt - Interacting Minds, Societies, Environments, Nicolaus Copernicus University in Toruń, Toruń, Poland

Hockett's Design features of language (e.g. Hockett, 1960; 1963) have proven highly influential within the field of linguistics in describing the differences between human language and animal communication. However, they have not remained without criticism over the many years since their inception. In particular, although frequently mentioned, they have "generally failed to motivate theoretical discussion or empirical tests" (Hauser, 1996: 47) in animal communication research. Focusing on language evolution, Wacewicz & Żywiczyński (2015) have argued that the design feature approach does not pay enough attention to the actual cognitive capacities underlying these features and the ecological and social settings in which they are used.

Despite these criticisms, Hockett's design features are still frequently appealed to by linguists. Here we argue that if linguistics wants to continue using them, this needs to be complemented by a decompositional approach that explicates a given design feature with regard to the cognitive and neuroscientific component parts that underlie the feature. Only by re-evaluating Hockett's design features from a cognitive and neuroscience perspective can linguistics be brought into a productive dialogue with animal communication research and language evolution research. Our proposal follows that of Wacewicz & Żywiczyński (2015) but extends it by adding a neuroscientific perspective where all of Hockett's design features are decomposed in a systematic fashion.

We illustrate this approach by focusing on the property of displacement, which suggests that, instead of being specific to human language, displacement is evolutionarily continuous. Displacement in language depends on episodic memory, which has been detected in nonhuman animals, and is supported by an ancient subcortical region—the hippocampus. This has important implications for an overall re-evaluation of Hockett's design features.

Displacement describes linguistic messages referring to events remote in both time and space (Hockett, 1963). Episodic memory, which enables one to revisit the past and imagine the future (Tulving, 2001), is necessary to achieve displacement. Comparative evidence reveals that although it might be rare, neither displacement nor episodic memory are unique to humans. As Hockett (1963) observes, the waggle dance of honeybees is always displaced (Root-Bernstein, 2010). In addition, chimpanzees have been shown to communicate about objects that are absent or displaced (Lyn et al. 2014; Bohn et al. 2015). Regarding episodic memory, "what-where-when" has been taken as a reference for identifying episodic-like memory in nonhuman animals. For example, food caching birds, rodents, and nonhuman primates have been shown to be able to accomplish episodic-like memory tasks (Clayton & Dickinson, 2010).

Neurological studies reveal that the hippocampus could play a key role in displacement. In humans, the hippocampus has been shown to be involved not only in episodic memory, but also in encoding spatial and temporal information (Howard & Eichenbaum, 2015). It is also assumed to store when, where and what information (Tsao et al., 2018). In birds, the behavior of food-caching engages the hippocampus. For instance, Black-capped chickadees and dark-eyed juncos whose hippocampus is lesioned exhibit memory impairment for location, which is a crucial part for food-caching (Colombo & Broadbent, 2000). It has also been reported that the seasonal change in hippocampal size is correlated with seasonal change in caching food in food-storing birds (Sherry & Hoshooley, 2010). In rodents, different subregions of the hippocampus are involved in subcomponents of episodic memory like recognition, temporal order and spatial memories (Aggleton & Pearce, 2001). The hippocampus has been shown to take part in object recognition and time delay between sample and test sessions (Cohen & Stackman, 2015). Specifically, the dorsal part is critical for temporal information processing, and the ventral part is crucial for temporal order of the spatial information (Hoge & Kesner, 2007; Howland et al., 2008). In nonhuman primates, the what-where association has been investigated by recording single neurons in macaques, and the results showed that the hippocampus is involved in separation and combination of representations of objects and the places where they are located (Roll et al., 2005). Zola et al. (2000) also found that hippocampal lesions affected object recognition memory.

As the case of displacement shows, a cognitively and neuroscientifically informed re-evaluation of Hockett's design features therefore offers the potential of turning them into a useful analytical device for comparisons of human language and animal communication.

## References

Aggleton, J. P., & Pearce, J. M. (2001). Neural systems underlying episodic memory: Insights from animal research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *356*(1413), 1467–1482.

Bohn, M., Call, J., & Tomasello, M. (2015). Communication about absent entities in great apes and human infants. *Cognition*, *145*, 63-72.

Clayton, N.S., & Dickinson, A. (2010). Mental Time Travel: Can Animals Recall the Past and Plan for the Future? In *Encyclopedia of Animal Behavior* (pp. 438–442). Elsevier.

Cohen, S. J., & Stackman, R. W. (2015). Assessing rodent hippocampal involvement in the novel object recognition task. A review. *Behavioural Brain Research*, *285*, 105–117.

Colombo, M., & Broadbent, N. (2000). Is the avian hippocampus a functional homologue of the mammalian hippocampus? *Neuroscience and Biobehavioral Reviews*, *24*(4), 465–484.

Hocket, C. D. (1960). The origin of speech. *Scientific American*, 203, 88-111.

Hockett, C.F. (1963). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals in language* (pp. 1-22). MIT Press

Hoge, J., & Kesner, R. P. (2007). Role of CA3 and CA1 subregions of the dorsal hippocampus on temporal processing of objects. *Neurobiology of Learning and Memory*, *88*(2), 225–231.

Howard, M. W., & Eichenbaum, H. (2015). Time and space in the hippocampus. *Brain research*, *1621*, 345-354.

Howland, J. G., Harrison, R. A., Hannesson, D. K., & Phillips, A. G. (2008). Ventral hippocampal involvement in temporal order, but not recognition, memory for spatial information. *Hippocampus*, *18*(3), 251–257.

Lyn, H., Russell, J. L., Leavens, D. A., Bard, K. A., Boysen, S. T., Schaeffer, J. A., & Hopkins, W. D. (2014). Apes communicate about absent and displaced objects: methodology matters. *Animal Cognition*, *17*(1), 85-94.

Rolls, E. T., Xiang, J., & Franco, L. (2005). Object, space, and object-space representations in the primate hippocampus. *Journal of Neurophysiology*, *94*(1), 833–844.

Root-Bernstein, M. (2010). Displacement activities during the honeybee transition from waggle dance to foraging. *Animal Behaviour*, *79*(4), 935–938.

Sherry, D. F., & Hoshooley, J. S. (2010). Seasonal hippocampal plasticity in food-storing birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1542), 933–943.

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M. B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, 561(7721), 57-62.

Tulving, E. (2001). Episodic memory and common sense: How far apart? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *356*(1413), 1505–1515.

Wacewicz, S., & Żywiczyński, P. (2015). Language evolution: why Hockett's design features are a non-starter. *Biosemiotics*, 8(1), 29-46.

Zola, S. M., Squire, L. R., Teng, E., Stefanacci, L., Buffalo, E. A., & Clark, R. E. (2000). Impaired recognition memory in monkeys after damage limited to the hippocampal region. *Journal of Neuroscience*, *20*(1), 451–463.

# AN EVOLUTIONARY PERSPECTIVE ON IM/POLITENESS AND THE PRAGMATIC CONCEPT OF FACE

MONIKA PLEYER*[1] and MICHAEL PLEYER[2,3]

*Corresponding Author: monika.pleyer@as.uni-heidelberg.de

[1] Department of English Literature and Linguistics, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany

[2] Center for Language Evolution Studies, Nicolaus Copernicus University in Toruń, Toruń, Poland

[3] University Center of Excellence IMSErt - Interacting Minds, Societies, Environments, Nicolaus Copernicus University in Toruń, Toruń, Poland

The concepts of im/politeness and face are central to pragmatic accounts of human communication. Im/politeness can be defined as behaviour designed to disarm or enhance aggression, whereas the notion of face broadly refers to a person's reputation in interaction. Given their centrality, both concepts should also play an important role in accounts of the evolutionary emergence of language and human communication. Here, we adopt an evolutionary perspective on im/politeness and face. We first describe the roles im/politeness and face play in human interaction, with a special focus on the subcomponents comprising the concept of face: self-image or self-worth, interpersonal affect, identity, and reputation. We then briefly sketch the evolutionary benefits of im/politeness and face, highlighting their role in co-operation and group management. Lastly, we discuss aspects of face shared with other primates, focussing especially on the process of reputation formation. We conclude that although reputation formation can be seen as an evolutionary foundation of the evolution of face, face management is uniquely expressed in language and human interaction.

## 1. Introduction

In this paper we adopt an evolutionary perspective on im/politeness and the pragmatic concept of face. Whereas language evolution research has increasingly focused on the importance of pragmatics (e.g., Scott-Phillips 2017), the concept of im/politeness and its relation to the pragmatic concept of face so far have hardly received any attention. Given its centrality to human interaction, an evolutionary perspective on im/politeness and face can prove highly valuable in shedding further light on the pragmatic foundations of language use and evolution (cf.

Pleyer & Pleyer 2016; Mühlenbernd et al. 2021). In this paper, we present such an evolutionary perspective, first focussing on theoretical and definitional work in pragmatics and im/politeness studies on the concepts of im/politeness and face. We will then take a look at the potential evolutionary benefits of im/politeness and face, and explore aspects of face that humans share with other primates. Overall, we will argue that many aspects of im/politeness and face are shared with other primates, but that they are uniquely expressed in human language.

## 2. Im/Politeness

Before we can discuss the evolution and evolutionary benefits of being im/polite, we must first define the object under study. In linguistics, there is a plethora of definitions of im/politeness (Culpeper 2011); for us, the following two are the most central. For Brown and Levinson (1987: 1), "politeness, like formal diplomatic protocol […] presupposes that potential for aggression as it seeks to disarm it, and makes possible communication between potentially aggressive parties." In other words, politeness relies on the interactants' wish to cooperate and to avoid conflict (see also Leech 1983). Impoliteness, on the other hand, seeks to create antipathy and non-cooperation (see also Kienpointner 1997): "Situated behaviours are viewed negatively – considered 'impolite' – when they conflict with how one expects them to be, how one wants them to be and/or how one thinks they ought to be." (Culpeper 2011: 23). This means that impoliteness includes a violation of (social) norms and/or expectations. It may trigger negative emotions, such as anger, shame, or sadness, and can have negative consequences, ranging from one or more participants feeling offence to the potential of physical altercations.

In our view, both politeness and impoliteness rely on the same shared cognitive mechanisms, and further, interpretations of behaviours as polite or impolite rely heavily on contextual factors. Thus, politeness and impoliteness should be seen as two end points of a scale rather than two distinct phenomena (see also Watts 2003). It is relevant for the evolution of im/politeness that both can be expressed using a potentially open-ended set of verbal and non-verbal behaviours. Polite verbal behaviours include e.g. complimenting, showing interest in H, or giving deference, whereas non-verbal politeness may include smiling or doing an action beneficial to H – in short, behaviours that make H feel good and thereby demonstrate the other's wish to cooperate and avoid conflict. Impoliteness, in turn, includes behaviours that make the hearer 'feel bad' (Culpeper 2011: 9), such as verbally insulting them, offering dispraise, showing disinterest or upsetting power relations; non-verbal behaviours may include shouting or frowning at H, or doing an action detrimental to their well-being. In short, many of these actions include an unmitigated potential for aggression. Many of these non-verbal expressions of im/politeness may well be shared with other non-human primates (Mason & Mendoza 1993).

## 3. Face – what it is, what it isn't

The pragmatic concept of face can be described as "the positive social value a person effectively claims for himself [sic] by the line others assume he has taken during a particular contact" (Goffman 1967: 5). This means that face is associated with personal, relational and social value, and is concerned with people's sense of worth, reputation, competence, etc. As such, it is understood as a dyadic, interpersonal phenomenon that is co-constituted in interaction (Arundale 2006). Face is thus a way of managing rapport, or self and other (Spencer-Oatey 2008). One must distinguish between individual face and group face, a form of face that highlights the concerns of the group over the individual (see e.g. many non-Western societies, Nwoye 1992). Face is further relevant for the expression of im/polite beliefs: polite behaviour seeks to maintain or enhance the hearer's face in that all potential face-threats must be redressed with politeness (see Brown & Levinson 1987), while impoliteness sets out to threaten it (Culpeper 1996).

In pragmatics, research has highlighted several concepts that are closely related to face, and that may be shared to a degree with other primates. These are self-image or self-worth, interpersonal affect, identity, and reputation.

One's self-image are ideas a person has about themselves and would like others to entertain, as well. It is distinct from face as one can have a self-image separate from interactions, but face can and is only relevant in interactions with others (see Goffman's 1967: 5 view of face as being "on loan from society"). Speakers may be able to modify their self-image (e.g. by dressing in a certain way) which might inform their face, but threats to face might, but do not automatically have to, incur loss of self-worth.

Interpersonal affect is similar to, but cannot be fully equated to face. For instance, the speaker's malign intent in a potentially face-threatening act can be overcome by H in pretending S had an originally supportive intent. Thus, in a group setting, H would not lose face, but the interpersonal relation between S and H might suffer (see O'Driscoll 2011: 20-21). Interpersonal affect requires quite a high level of social awareness, as well as knowledge about the distinction of individual and group face.

Identity and face are closely interconnected aspects. Some researchers in essence equate the two concepts; for instance, for Sifianou (2011: 42), face "can be seen as a positive social image akin to identity." Zimman (2018: 178-79) describes identity creation as an intersubjective construction together with other interactants, which bears strong similarities to Goffman's (1967) understanding of face. Other stances see a clear distinction between these concepts in that face is bestowed upon ego by society post-factum, whereas identity is also bestowed upon ego by ego (Spencer-Oatey 2007). Arundale (2006: 202) echoes this view, and further understands identity as a phenomenon much broader than face. In our view, identity then requires a knowledge of the self as an individual, and as the individual having some control over how self is presented to others. Some aspects of this phenomenon are likely shared with a small subset of animals, e.g. those that can recognise themselves in mirrors, such as cetaceans, great apes, magpies, crows, and Asian elephants (De Waal 2019). Some animals, such as chimpanzees, might even have a more developed concept of self extending across time and space, as indicated in their ability to recognise themselves in delayed video footage (Hirata et al. 2017).However, the concept closest to face that is most central for our understanding of primate interactions seems to be that of reputation. Reputation can be described as "a universal currency for human social interactions" (Milinski 2016), and signals the likely behaviour of an individual (see also Mühlenbernd et al. 2021). Thus, there is quite a strong overlap with Goffman's (1967: 5) 'line' that others assume we take in a given interaction. As the concept of reputation is used both in pragmatics and evolutionary studies, it seems an ideal concept for interdisciplinary research

## 4. The evolutionary benefits of im/politeness and face

In our view, the aspects of politeness most relevant to early humans are that of cooperation and avoiding conflicts. Behaving in accordance with the group face, and avoiding conflicts secures the group's cohesion, and as such, can be beneficial to both the group's survival and, more importantly, the survival of individual members within the group. This can be strengthened with emerging conventions of how to do things (with words), i.e. shared sets of pragmatic norms which, if followed, demonstrate the individual's belonging to a group. In this way, individuals who adhere to these emerging conventions within such groups accrue benefits to their own evolutionary fitness. We can observe similar behaviours in non-human apes. For example, research has shown that baboons, who live in groups up to 100 individuals, show an awareness for social norms and power structures (Bergman et al. 2003).

However, there are also evolutionary benefits to impoliteness, as it can also be used as a method of in-group management (Kienpointner 1997). As such, impoliteness secures group norms by excluding those individuals who are unwilling or unable to follow them. This generates a group of like-minded individuals who share the same conventions, which, as stated above, may aid in survival. It further enhances group stability in generating an ‘us’ vs ‘them’ mentality, which may aid in e.g. resource distribution. Just as for politeness, an individual who successfully manages their belonging to and status within a stable in-group will enhance their own evolutionary fitness.

## 5. Aspects of face shared with other primates

Many non-human animals also exhibit complex social cognition. This holds especially for non-human primates. Chimpanzees, for example, understand psychological states like seeing, knowledge and ignorance, know that others make inferences, and understand others’ perceptions and intentions (e.g., Bettle & Rosati 2020). These capacities have the potential to shed light on the evolutionary origins of the human interactive concept of face and the question of which aspects of face are shared with other primates.

As mentioned in Section 3, one aspect of face that is also crucial to primate and many non-human animal interactions is that of reputation. From a comparative psychological perspective, reputation can be defined as “knowledge of an individual’s typical behaviour in a situation based on their past behaviour” (Russel et al. 2008). Acquiring knowledge about an individual’s reputation is known as reputation formation (Herrmann et al. 2013). There are three principal ways of reputation formation (Smith & Harper 2003): through direct experience, through observation of other individuals interacting with each other, and through gossip, that is, receiving third-party information about somebody’s behaviour. In humans, all three sources contribute to reputation formation. Other animals, however, seem not to make use of communicative signals that inform others of the reputation of a third party. However, for some highly social non-human animal species, reputation formation might not only be based on direct interactions, but also on observations of others’ interactions. For example, Subiaul et al. (2008) found that chimpanzees preferred to interact with individuals that they observed to be ‘generous’ to others versus those that they observed to be ‘selfish’. Similarly, Russel et al. (2008) found that chimpanzees preferred individuals they observed to exhibit ‘nice’ behaviour to those that exhibited ‘nasty’ behaviour. However, they found no such preference for gorillas, bonobos, and orangutans. Herrmann et al. (2013) found that both chimpanzees and orangutans preferred ‘nice’ to ‘mean’ individuals, with bonobos showing no such preference. In all three experimental designs, ‘generous’/’nice’ vs ‘selfish’/’nasty’/’mean’ was operationalised as whether an individual would share food with somebody else who begged for food. This research therefore indicates that in addition to direct reputation formation through interactions with conspecifics, indirect reputation judgement is present to at least some degree in some of the other great apes.

However, the potential mechanisms involved in this process are still being debated. Earley (2010) describes this process as ‘social eavesdropping’, where bystanders attend to, and use, information emitted by signallers; Russel et al. (2008) refer to a more specific process of ‘image scoring’, where person A monitors the giving behaviour of person B towards person C. Finally, Abdai & Miklòsi (2016) speak of a general process of social evaluation, which they describe as “a mental process that leverages the preference toward prosocial partners (positivity bias) against the avoidance of antisocial individuals (negativity bias) in a cooperative context.” Given that social evaluation has also been observed in dogs and cleaner-client reef fish interactions (Abdai & Miklòsi 2016), it is possible that a general prosocial vs antisocial bias provides an evolutionary ancient foundation for the evolution of the human concept of face.

There is, however, one aspect related to face and reputation that seems not to be present in other primates to the same degree, namely that of reputation management. Humans seem to be uniquely aware and concerned with how others see them, and how their behaviour is evaluated collectively (Tomasello 2019). For example, Engelmann et al. (2012) showed that 5-year-old children steal significantly less and help significantly more in a scenario where they are observed by others, as opposed to a scenario where they were unobserved. The same was not found for chimpanzees, where there was no significant difference in stealing behaviour depending on whether they were observed or not. Actively trying to manage one’s reputation therefore seems to be a crucial aspect of human face.

## 6. Conclusion

In this paper, we have addressed some important aspects of the evolution of im/politeness and face that have so far not received enough attention in pragmatic accounts of the evolution of language and human communication. Both im/politeness and face have evolutionary benefits, e.g. for group and resource management. Both humans and non-humans have complex sociocognitive capacities, but humans seem to be especially good at cultural learning, including im/politeness strategies, and group-mindedness.

Many components of face seem to be shared with other highly social species, especially other primates. As such, mapping animal behaviours on aspects of face can help develop an evolutionary perspective on its emergence in humans (cf. Lim & Bowers 1991). However, the explicit management and negotiation of aspects of face seems to be uniquely expressed in human interaction, and especially so through the use of language.

## References

Abdai, J., & Miklósi, Á. (2016). The origin of social evaluation, social eavesdropping, reputation formation, image scoring or what you will. *Frontiers in Psychology*, 7, 1772.

Arundale, R. (2006). Face as relational and interactional: A communication framework for research on face, facework and politeness. *Journal of Politeness Research*, *2/2,* 193–216.

Bettle, R., & Rosati, A. G. (2020). The primate origins of human social cognition. *Language Learning and Development*, 1-32.

Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science, 302(5648),* 1234–1236.

Brown, P., & Levinson, S.C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Culpeper, J. (2011). *Impoliteness. Using language to cause offence*. Cambridge: Cambridge University Press.

Culpeper, J. (1996). Towards an anatomy of impoliteness. *Journal of Pragmatics* 25, 349-367.

de Waal, F. B. (2019). Fish, mirrors, and a gradualist perspective on self-awareness. *PLoS Biology*, *17*(2), e3000112.

Earley, R. L. (2010). Social eavesdropping and the evolution of conditional cooperation and cheating strategies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553), 2675-2686

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS One*, 7(10), e48433.

Goffman, E. (1967). *Interaction ritual. Essays on face-to-face behavior*. Garden City, NY: Doubleday Anchor.

Herrmann, E., Keupp, S., Hare, B., Vaish, A., & Tomasello, M. (2013). Direct and indirect reputation formation in nonhuman great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus) and human children (Homo sapiens). *Journal of Comparative Psychology,* 127(1), 63.

Hirata, S., Fuwa, K., & Myowa, M. (2017). Chimpanzees recognize their own delayed self-image. *Royal Society Open Science*, *4*(8), 1-9.

Kienpointner, M. (1997). Varieties of rudeness: Types and functions of impolite utterances. *Functions of Language 4, (2),* 251–87.

Leech, G.N. 1983. *Principles of pragmatics*. London: Longman.

Lim, T.-S., and Bowers, J.W. (1991). Facework: solidarity, approbation, and tact. *Human Communication Research*, 17, 415–450.

Mason, W. A., & Mendoza, S. P. (Eds.). (1993). *Primate social conflict*. SUNY Press.

Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1687), 20150100. https://doi.org/10.1098/rstb.2015.0100

Mühlenbernd, R., Wacewicz, S., & Żywiczyński, P. (2021). Politeness and reputation in cultural evolution. *Linguistics and Philosophy, 44*(6), 1181-1213.

Nwoye, O.G. (1992). Linguistic politeness and socio-cultural variations of the notion of face. *Journal of Pragmatics*, 18/4, 309–328.

O'Driscoll, J. (2011). Some issues with the concept of face: When, what, how, and how much? In F. Bargiela-Chiappini & D. Z. Kádár (Eds.). *Politeness across cultures* (pp. 17-41). Basingstoke: Palgrave Macmillan.

Pleyer, M., & Pleyer, M. (2016). The Evolution of Im/politeness. In S. G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Feher, & T. Verhoef (Eds.), *The Evolution of Language. Proceedings of the 11th International Conference on the Evolution of Language (EVOLANG 11).*

Russell, Y. I., Call, J., & Dunbar, R. I. (2008). Image scoring in great apes. Behavioural Processes, 78(1), 108-111.

Scott-Phillips, T. C. (2017). Pragmatics and the aims of language evolution. Psychonomic Bulletin & Review, 24(1), 186-189.

Sifianou, M. (2011). On the concept of face and politeness. In F. Bargiela-Chiappini & D. Z. Kádár (Eds.), *Politeness across cultures* (pp. 42-58). Basingstoke: Palgrave Macmillan.

Smith, J. M., & Harper, D. (2003). *Animal signals*. Oxford University Press.

Spencer-Oatey, H. (2008). Face, impoliteness and rapport. In Spencer-Oatey, Helen (Ed.), *Culturally speaking: Managing rapport through talk across cultures* (pp. 11-47). 2nd edn. London: Continuum.

Spencer-Oatey, H. (2007). Identity, face and (im)politeness. *Journal of Pragmatics, 39(4)*, 635-638.

Subiaul, F., Vonk, J., Okamoto-Barth, S., & Barth, J. (2008). Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers. *Animal cognition, 11*(4), 611-623.

Tomasello, M. (2019). *Becoming human*. Harvard University Press.

Watts. R.J. (2003). *Politeness*. Cambridge: Cambridge University Press.

Zimman. L. (2018). Pronouns and possibilities. Transgender language activism and reform. In N. Avineri, L.R. Graham, E.J. Johnson, R.C. Riner and J. Rosa (Eds.), *Language and social justice in practice* (pp. 176-183). New York; London: Routledge.

# THE ‘MUSICAL’ ORIGIN OF LANGUAGE GRAMMAR AS A MILESTONE FOR THE DEVELOPMENT OF A COMPLEX CONCEPTUAL MIND

PIOTR PODLIPNIAK[1]

[*]Corresponding Author: podlip@amu.edu.pl
[1]Department of Musicology, Adam Mickiewicz University in Poznań, Poland

## 1. Abstract

Speech is composed of two elements: 1) affective prosody (Brown, 2017) and 2) articulate vocalizations which are claimed to be processed by two different neurological pathways (Ackermann et al., 2014). At the psychological level, affective prosody is based on continuous changes of parameters such as pitch and loudness, whereas articulate vocalizations consist of perceptually discrete entities such as phonemes and words that enable combinatoriality (Hilliard & White, 2009). While articulate vocalizations transmit meaning by means of the exchange of mental concepts, affective prosody acts as a source of emotional induction. Moreover, compared to articulate vocalizations, affective prosody is less susceptible to volitional control (Ackermann et al., 2014). As language grammar seems to be a framework for the exchange of complex conceptual meanings, its origin is usually claimed to be related to articulate vocalizations and the conceptual mind (Ginsburg & Jablonka, 2019). However, the crucial features of language grammar – hierarchy and recursion – are also present in music (Pinker & Jackendoff, 2005; Rohrmeier et al., 2015) which is devoid of propositional semantics (Lerdahl, 2013), and which shares many traits with affective prosody. The main aim of this theory-based proposal is to present a scenario in which language grammar might have originated from primary ‘felt’ sound relations used to induce emotions, rather than from articulate symbols. From this perspective, the mechanisms that enabled proto-musical forms of rhythm and pitch hierarchies could have been exapted to serve a hierarchical framework for sound (and gestural) symbols in order to enable the exchange of

information about increasing social complexity. This exaptation could have led to the emergence of language grammar and a complex conceptual mind.

In order to trace the sequence of events that led to the evolution of language grammar, a comparison between speech and music is proposed. For instance, the presence of the elements of affective prosody in many mammalian vocalizations, as well as in all forms of human vocal expressions (Ross et al., 2009), can lead us to assume that speech prosody consists of features that evolved long before the emergence of language. In addition, propositional meaning in the form of referential concepts seems to be much older than *Homo sapiens*. These concepts could have probably been externalized by means of both sounds and gestures long before the appearance of hominins. After all, the simple forms of such externalizations are still observed in apes (Kalan et al., 2015).

It has been proposed that language could have evolved as a result of a merger between two functionally different communicative protolanguages: one designed to express external meaning and another designed to express internal meaning (Podlipniak, 2022). If this scenario is true, the human conceptual mind owes its complexity to the evolution of both of these communicative systems. Although prosody is usually perceived in a continuous way it can also be experienced, under some conditions, as being composed of hierarchically organized discrete pitch and rhythm units (Deutsch et al., 2008), which suggests that hierarchization is achievable without semantics. Moreover, as simple temporal hierarchies have been observed in infant-directed speech and song (Falk & Kello, 2017), as well as in non-human vocalizations (Ravignani et al., 2019) which are devoid of propositional meaning, it is reasonable to assume that hierarchical communication has its roots in the affective prosody. This view is supported by the fact that in the least grammatically complex natural languages the grammatical relations depend mainly on prosody (Gil, 2005; Jackendoff & Wittenberg, 2014; Sandler, 2017). From this perspective, the appearance of a complex semantic compositionality was possible thanks to the exploitation of the previously existing mechanisms that enabled hierarchical ordering of affective sounds in the domain of symbolic articulate vocalizations. In other words, the emergence of language grammar could have been the result of the exaptation of cognitive mechanisms used in preconceptual proto-musical expressions into the creation and externalization of hierarchical relations between concepts and words.

## References

Ackermann, H., Hage, S. R., Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *37*(06), 529–546.

Brown, S. (2017). A Joint Prosodic Origin of Language and Music. *Frontiers in Psychology, 8,* 1894.

Deutsch, D., Lapidis, R., Henthorn, T. (2008). The speech-to-song illusion. *The Journal of the Acoustical Society of America*, *124*(4), 2471.

Falk, S., Kello, C. T. (2017). Hierarchical organization in the temporal structure of infant-direct speech and song. *Cognition*, *163*, 80–86.

Gil, D. (2005). Word Order Without Syntactic Categories: How Riau Indonesian Does It. In A. Carnie, S. A. Dooley, H. Harley (Eds.), *Verb first: on the syntax of verb-initial languages* (pp. 243–263).

Ginsburg, S., Jablonka, E. (2019). *The Evolution of the Sensitive Soul - Learning and the Origins of Consciousness*. The MIT Press.

Hilliard, A., White, S. A. (2009). Possible Precursors of Syntactic Components in Other Species. In D. Bickerton, E. Szathmáry (Eds.), *Biological Foundations and Origin of Syntax* (pp. 161–183). The MIT Press.

Jackendoff, R., Wittenberg, E. (2014). What you can say without syntax: a hierarchy of grammatical complexity. In F. J. Newmeyer L. B. Preston (Eds.), *Measuring Grammatical Complexity* (pp. 65–82).

Kalan, A. K., Mundry, R., Boesch, C. (2015). Wild chimpanzees modify food call structure with respect to tree size for a particular fruit species. *Animal Behaviour*, *101*, 1–9.

Lerdahl, F. (2013). Musical syntax and its relation to linguistic syntax. In M. A. Arbib (Ed.), *Language, Music, and the Brain* (pp. 257–272). MIT Press.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What’s special about it? *Cognition*, *95*(2), 201–236.

Podlipniak, P. (2022). Pitch syntax as part of an ancient protolanguage. *Lingua*, *271*, 103238a. https://doi.org/10.1016/J.LINGUA.2021.103238

Ravignani, A., Dalla Bella, S., Falk, S., Kello, C. T., Noriega, F., Kotz, S. A. (2019). Rhythm in speech and animal vocalizations: a cross-species perspective. *Annals of the New York Academy of Sciences*, nyas.14166.

Rohrmeier, M., Zuidema, W., Wiggins, G. A., & Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140097.

Ross, M. D., Owren, M., Zimmermann, E. (2009). Reconstructing the Evolution of Laughter in Great Apes and Humans. *Current Biology*, *19*, 1106–11.

Sandler, W. (2017). What comes first in language emergence? In N. J. Enfield (Ed.), *Dependencies in language: on the causal ontology of linguistic systems* (pp. 63–83).

# EVOLANG DATABASE: A COMPREHENSIVE SOURCE OF INFORMATION ON THE SCIENCE OF LANGUAGE EVOLUTION COMMUNITY

ALEKSANDRA PONIEWIERSKA*[1], ALEKSANDRA SZCZEPAŃSKA[1], YEN YING NG[1], MARTA SIBIERSKA[1], MAREK PLACIŃSKI[1], PRZEMYSŁAW ŻYWICZYŃSKI[1] and SŁAWOMIR WACEWICZ[1]

*Corresponding Author: maestro.aleksandra@gmail.com
[1] Center for Language Evolution Studies, Nicolaus Copernicus University in Toruń, Toruń, Poland

## 1. Introduction

In recent years, language evolution has further consolidated its identity as a distinct field of research (see e.g. Dediu & de Boer 2016; Kirby 2017; cf. the "Science of Language Evolution", Żywiczyński 2018) and has become the object of interest of metascientific analyses (see e.g. Bergmann and Dale 2016; Roberts et al. 2020). To contribute to this trend, we created a comprehensive database of information on the Evolang conference series (available at http://sle.cles.umk.pl/), as representative of the language evolution academic community. In what follows, we discuss the applicability of this database for scientometric research and present a case study of the academic profiles of those researchers involved in Evolang who self-identify with interests and disciplines most commonly associated with language evolution.

## 2. Data collection

The database contains information on Evolang abstracts (n = 1085) and Evolang contributors (n = 1401) over the period of 9 most recent iterations of the conference (2004-2020). The database was created in four steps: (1) A list of contributors was compiled based on automatic text mining of the Evolang books of abstracts. (2) Basic bibliographic information was collected from the proceedings volumes through automatic data collection, which allowed us to identify the author profiles in Google Scholar (GS). Despite certain drawbacks, most importantly incomplete coverage of author profiles, GS has already been acknowledged as a useful tool in scientometric analyses (e.g. Gusenbauer 2019), and it proves particularly informative in studying self-reported data, such as one's discipline or research interests (see below). (3) Each author's profile was supplemented with data from GS (automatic data collection). The data collected includes self-reported affiliation and research interests, the year of the author's first publication shown in GS, as well as scientometric information such as h-index, i10-index, and citation count. (4) All the data collected was verified by three independent experts (manual data verification).

## 3. Database design

The database is organized into four tables: (1) "authors", including all Evolang contributors, (2) "publications", including bibliometric information on all Evolang abstracts, (3) "references", including full references used in the abstracts, and (4) "texts", including all the abstracts themselves. Tables (1), (2), and (3) are available in the user interface view of the database (see: http://sle.cles.umk.pl/). The data is fully interoperable and reusable, and in the future extended implementations it will be supplemented with additional information, such as on the geographical distribution of the authors.

| *authors* | | *publications* | |
|---|---|---|---|
| *field* | *source* | *field* | *source* |
| affiliation | GS | author(s): first and last names, affiliations | B |
| country | GS | excerpt | B |
| ORCID | M | references: all references used in the abstract | B |
| discipline | GS | citations: all citations used in the abstract | B |
| first Evolang contribution | B | type of document: short/long abstract | B |

| | | | |
|---|---|---|---|
| number of Evolang contributions | B | type of presentation: regular talk/plenary | B |
| first scientific publication | GS | bibliographic information | B |
| h-index (5-year and overall) | GS | | |
| i10 index (5-year and overall) | GS | | |
| citations | GS | | |

Table 1. A summary of the fields available in tables (1) and (2) of the database, and the relevant sources of data, where M stands for manual, GS stands for Google Scholar, and B for books of abstracts. Table (3) corresponds to the "publications/references" field; table (4) is partly available through "excerpts".

## 4. Analyses

To exemplify the type of questions that can be addressed with the current implementation of the database, we present an analysis of the self-identification of the Evolang contributors with a number of labels most often used to denote language evolution as a field of research. 642 authors in our database had identifiable GS profiles, with a total of 916 tokens of associated discipline/field labels (self-reported in GS as "Areas of interest"). Here, we looked into five labels: "language evolution", "evolutionary linguistics", "evolution of language", "language origin(s)" / "origin of language", and "biolinguistics" (see Table 2 below). This analysis helps us explore the subtle connotational differences between those five labels that are taken to be near-synonymous or overlapping. Of those five, "language evolution" is by far the most frequently used label by Evolang contributors, whereas "evolutionary linguistics" appears to be preferentially used by more senior and more accomplished researchers, oriented towards more computational research, and more strongly involved in the Evolang conference (of the total of four existing Google Scholar profiles with this label, three belonged to Evolang contributors). Perhaps surprisingly, only three Evolang contributors report "biolinguistics" as their area of research in their GS profile.

| | *language evolution* | *evolutionary linguistics* | *evolution of language* | *language origin(s), origin of language* | *biolinguistics* |
|---|---|---|---|---|---|
| 1. Number of authors | 69 | 15 | 21 | 4 | 3 |
| 2. Proportion Evolang to GS | .570 | .750 | .636 | .333 | 0.125 |
| 3. Other most common labels | cultural evolution, cognitive science, linguistics | computational linguistics, cognitive science | cognitive linguistics, comparative cognition | language evolution, linguistics | NA* |
| 4. Per cent long papers | 17.85 | 31.67 | 23.15 | 30 | 33.33 |
| 5. Seniority (years since first GS publication) | 15.94 | 26.27 | 18.63 | 16 | 16.33 |
| 6. Mean citation count | 1494.66 | 4825.4 | 2436.74 | 464 | 899 |

Table 2. A summary of the analysis of selected Google Scholar labels (self-declared "areas of interest") of Evolang contributors. Row 1 states the total number of Evolang authors for a given GS label. Row 2 states the share of Evolang authors among all GS profiles for that label. Row 3 states the other GS labels most commonly declared by Evolang authors with a given label. Row 4 states the proportion of papers (6-8 pages) to abstracts (2 pages) submitted by Evolang authors; we assumed the submission type to reflect the distinction between review / theoretical papers (typically longer) and reports of empirical results (typically shorter). Row 5 reports author seniority, operationalised as the time of active publishing, counting from the first publication listed on GS. Row 6 reports GS citation count.

## 5. Conclusions

As the field of language evolution research has rapidly grown and now constitutes to develop as a broad and highly interdisciplinary field, researchers have expressed the need to seek new methods to systematize and explore the scientific production in this field (Bergmann & Dale, 2016; Roberts et al. 2020). We believe that the database reported in this contribution provides a valuable tool and resource to this end, with a range of theoretical and practical applications. In future work, we plan enriching the database with information from complementary resources, which would further extend the range of possible analyses. One example is the inclusion of GS citation data on individual publications, which will make it possible to identify the classic references (i.e. the publications most often – and most interdisciplinarily widely – cited in the Evolang proceedings), and address questions on the nature of the contribution of the different disciplines, research centers and author networks and its dynamics across the successive iterations of the Evolang conference. Another interesting direction is integrating our database with thematically related databases – in particular, the CHIELD database (Roberts et al. 2020), which contains entries for a large number of Evolang articles, with manually coded information such as on the methods used in the reported studies (e.g. "experiment", "observation") and the stages of language emergence to which they refer (e.g. "biological evolution", "language change").

## Acknowledgements

This research was supported by the Polish National Science Centre under grant agreement UMO-2019/34/E/HS2/00248.

## References

Bergmann, T., and Dale, R. (2016). A Scientometric Analysis Of Evolang: Intersections and Authorships. In S. G. Roberts et al. (eds) *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11).* New Orleans: Evolang Scientific Committee. Availa.

Dediu, D., & Boer, B.G. (2016). Language evolution needs its own journal. *Journal of Language Evolution, 1,* 1-6.

Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, *118(1),* 177-214.

Kirby, S. 2017. Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review 24(1)*, 118-137.

Roberts, S. et al. (2020). CHIELD: The Causal Hypotheses in Evolutionary Linguistics Database, *Journal of Language Evolution, 5(2)*, 101-120, DOI: 10.1093/jole/lzaa001.

Żywiczyński, P. (2018). *Language Origins: From mythology to science.* Berlin: Peter Lang.

# ELEPHANTS AS A NEW ANIMAL MODEL FOR STUDYING THE EVOLUTION OF LANGUAGE AS A RESULT OF SELF-DOMESTICATION

LIMOR RAVIV*[1,2,3], SARAH L. JACOBSON[4,5], JOSHUA M. PLOTNIK[4,5], JACOB BOWMAN[6], VINCENT LYNCH[6] AND ANTONIO BENÍTEZ-BURRACO[7]

*Corresponding Author: limor.raviv@mail.huji.ac.il
[1] LEADS group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] cSCAN, University of Glasgow, UK
[3] Artificial Intelligence Lab, Vrije Universiteit Brussels, Belgium
[4] The Graduate Center, City University of New York, USA
[5] Hunter College, City University of New York, USA
[6] Department of Biological Sciences, University at Buffalo SUNY, USA
7 University of Seville, Spain

## 1. Self-domestication and the evolution of language

Humans are unique in their sophisticated culture and societal structures, their complex languages, and their extensive tool use. But why did humans, and no other animal, develop these sets of complex traits? According to a prominent theory which is gaining more and more attraction, namely, the Human Self-Domestication hypothesis (HSD), this unique set of traits may be the result of an evolutionary process of self-induced domestication, in which humans have evolved to be more cooperative and prosocial as a result of environmental pressures in the middle and late Paleolithic (Hare, 2017; Benítez-Burraco, Clay & Kempe, 2020). According to the HSD hypothesis, self-domestication has resulted in less aggressive individuals who were more prone to interact with others (e.g., with kins and in-group members, but also with strangers), leading to increased social contacts and community structure, as well as more sophisticated forms of explicit teaching, learning, and exploration. Crucially, these properties may have ultimately given rise to the specific niche that enabled many of humans' distinctive traits, including our sophisticated linguistic abilities. The potential explanatory link between self-domestication and language evolution is gaining momentum in the field, with HDS invoked to explain many of the biological and cognitive changes that underlie the process of the cultural evolution of language (Thomas & Kirby, 2018; Progovac & Benítez-Burraco, 2019), as well as many critical features of modern languages such as pragmatics and turn-taking (Benítez-Burraco, Ferretti & Progovac, 2021), child-directed speech (Benítez-Burraco & Kempe, 2018), grammar sophistication and innovation (Langley, Benítez-Burraco & Kempe, 2019), , and cross-linguistic variability (Benítez-Burraco, 2020; Benítez-Burraco & Progovac, 2020).

While HSD is a compelling theory, it is hard to test. Most notably, the only other species that has been argued to be self-domesticated besides humans is Bonobos (Hare et al, 2012; Hare 2017; Kovalskas et al., 2021), resulting in a very narrow scope for investigating this theory beyond the primate family.

## 2. Are Elephants also self-domesticated?

Our work poses the novel question of whether Elephants may have also undergone a process of self-domestication, similar to humans and bonbons. If true, Elephants may serve as a new animal model for studying the potential impact and triggers of a self-domestication, and specifically the cultural evolution of language. Since the most recent common ancestor of humans and Elephants is likely the most recent common ancestor of all placental mammals, comparing the process of self-domestication and its outcomes in these species can lead to important insights about convergent evolution beyond the primate taxa, and open the door for exciting new research and better understanding of this process and its implications to communicative complexity in the broad sense.

We first motivate our hypothesis of Elephants as a self-domesticated species with an extensive cross-species comparison of relevant cognitive, behavioral, and physiological similarities between Elephants and the two other species that have been put forth as self-domesticated (i.e., humans and bonobos). This cross-species comparison demonstrates that Elephants indeed exhibit many of the features associated with self-domestication, including reduced aggression, increased prosocial behavior, extended juvenile period, increased play behavior, socially regulated cortisol levels, and complex vocal behavior. With respect to linguistic abilities, Elephants rely on a rich multimodal communication system that includes an extensive repertoire of vocal, visual, tactile and chemical signals to mediate the intricate teamwork displayed by members of an Elephant family (Jacobson & Plotnik, 2020). Their vocal signals have been shown to encompass specific meanings and intents shared with other members of the herd (e.g., Kahl & Armstrong, 2000; Poole & Granli, 2003; Wierucka, Henley & Mumby, 2021), and it shows a degree of intra-specific variation within and between individuals and across different herds - which is seen as an important feature of human language (e.g., Evans & Levinson, 2009). Next, we present novel genetic evidence to support our proposal and expose the positive selection of candidate genes for domestication in African Elephants. These novel analyses provide empirical evidence for our hypothesis: they show that genes positively selected in African Elephants are enriched in pathways that are involved in domestication, and that several candidate genes associated with domestication have been positively selected in African Elephants. We also discuss several explanations for what may have triggered a self-domestication process in the Elephant lineage, and discuss the potential implications of our theory to the field of language evolution.

## References

Benítez-Burraco, A. (2020). The Self-Domestication Hypothesis of Language Evolution. *Paradigmi* [Preprint].

Benítez-Burraco, A., Clay, Z. & Kempe, V. (2020). Self-Domestication and Human Evolution. *Frontiers in Psychology*, 11.

Benítez-Burraco, A., Ferretti, F. & Progovac, L. (2021) 'Human Self-Domestication and the Evolution of Pragmatics', *Cognitive Science*, 45(6), e12987.

Benítez-Burraco, A. & Kempe, V. (2018). The emergence of modern languages: Has human self-domestication optimized language transmission?. *Frontiers in Psychology*, 9.

Benítez-Burraco, A. & Progovac, L. (2020). A four-stage model for language evolution under the effects of human self-domestication. *Language & Communication*, 73, 1–17.

Evans, N. & Levinson, S.C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(5), 429–448.

Hare, B. (2017). Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality. *Annual Review of Psychology*, 68(1), 155–186.

Hare, B., Wobber, V. & Wrangham, R. (2012). The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour*, 83(3), 573–585.

Jacobson, S.L. & Plotnik, J.M. (2020). Elephant Cognition. In *Zoo Animal Learning and Training*. John Wiley & Sons, Ltd, 191–196.

Kahl, M.P. & Armstrong, B.D. (2000). Visual and tactile displays in African elephants, Loxodonta africana: a progress report. *Elephant,* 2(4), 19–21.

Kovalaskas, S., Rilling, J.K. & Lindo, J. (2021). Comparative analyses of the Pan lineage reveal selection on gene pathways associated with diet and sociality in bonobos. *Genes, Brain and Behavior*, 20(3), e12715.

Langley, M.C., Benítez-Burraco, A. & Kempe, V. (2019). Playing with language, creating complexity: Has play contributed to the evolution of complex language?. *Evolutionary Anthropology: Issues, News, and Reviews*, 29(1).

Poole, J.H. & Granli, P. (2011). Signals, Gestures, and Behavior of African Elephants. In *The Amboseli Elephants: A Long-Term Perspective on a Long-Lived Mammal*. University of Chicago Press, pp. 109–124.

Progovac, L. & Benítez-Burraco, A. (2019). From Physical Aggression to Verbal Behavior: Language Evolution and Self-Domestication Feedback Loop. *Frontiers in Psychology*, 10.

Thomas, J. & Kirby, S. (2018). Self domestication and the evolution of language. *Biology & Philosophy*, 33(1).

Wierucka, K., Henley, M.D. & Mumby, H.S. (2021). Acoustic cues to individuality in wild male adult African savannah elephants (Loxodonta africana). *PeerJ*, 9, e10736.

# THE POWER OF IDEAS FROM INFORMATION THEORY FOR STUDYING ANIMALS' NATURAL COMMUNICATION

ZHANNA REZNIKOVA *[1,2]

*zhanna@reznikova.net
[1]Institute of Systematics and Ecology of Animals, Novosibirsk, Russia
[2]V. Zelman Institute for Medicine and Psychology, Novosibirsk State University, Russia

Evolutionary linguistics has emerged as an attempt to answer questions concerning language evolution and human nature based on an interdisciplinary collaboration (Bickerton, 1990; Tomasello, 2008; Gong et al., 2014). The fundamental question about the uniqueness of human language is still far from being understood (Wacewicz & Żywiczyński, 2015). One of the possible paths forward suggested by Hauser et al. (2014) includes observations and experiments of naturally communicating animals and experiments assessing animals' computational and perceptual capacities, focusing on abilities necessary for human language processing. Here I compare three main experimental paradigms in studying animal language behavior and highlight the new dimension in this field provided by information theory.

The first approach is aimed at direct decoding the function and meaning of animal signals, which is a notoriously tricky problem. Two types of natural messages decoded up to the present concern the symbolic honeybee "dance language" (for a review, see Kohl et al., 2020) and fragments of acoustic communications in several species such as monkeys (review in Fischer, 2020) and dolphins (King & Janik, 2013). Girard-Buttoz et al. (2022) recently revealed a highly versatile vocal system in chimpanzees; however, larger datasets are still needed to prove animals' capacity to produce flexible vocal sequences that support numerous differentiated meanings. The second approach based on artificial intermediary languages uncovered significant "linguistic" and cognitive potential in some species that contradicts limitations in understanding their natural communications. The third approach applies ideas of Shannon entropy and Kolmogorov complexity. The main point is to study natural communications and evaluate their capabilities by measuring information transmission rates (details in: Ryabko & Reznikova, 2009; Reznikova & Ryabko, 2011). Without studying the nature of the signals used for communication, this approach provides a new dimension for understanding the essentials of animal communication systems.

Ants appeared to be even better candidates for studying general communication rules than the iconic honeybee. Among about 15 000 ant species, the majority of which display relatively simple modes of communication, a few *Formica rufa* species possess the leader-scouting system based on a consistent personal difference between scouting and foraging individuals (Reznikova, 2021). The idea of experiments is that the experimenters know the quantity of information (in bits; Shannon, 1948) to be transferred, which corresponds to the number of turns in a "binary tree" maze towards a "leaf" containing syrup. There were no cues that could help the ants find the food (including olfactory ones) except the information contacts with scouts. The duration of the contacts between the scouts and foragers appeared to be $ai + b$, where $i$ is the number of turns, $a$ is the time duration required for transmitting one bit of information, and $b$ is an introduced constant. The rate of information transmission was about 1 minute per bit in ants, which is at least 10 times smaller than in humans. Ants appeared to be able to grasp regularities in the "text" (sequence of turns, see below) to be transferred and use them to "compress" and thus optimize their messages. This series of experiments was inspired by the concept of Kolmogorov complexity (Kolmogorov, 1965) applied to words (or "texts") composed of the letters of an alphabet, for example, consisting of two letters: L (left) and R (right) corresponding to a sequence of turns in a "binary tree". Informally, the Kolmogorov complexity of a word (and its uncertainty) equates to its most concise description. For example, the word "LRLRLRLR" can be represented as "4LR", while the "random" word of shorter length "LRRLRL" probably cannot be expressed more concisely, and this is more complex. The hypothesis being tested was $H_0$, that is, the time for transmission of information by the scout does not depend on the complexity of the "text". The alternative hypothesis was $H_1$ that this time actually depends on the complexity of the "text". The hypothesis $H_0$ was rejected ($P = 0.01$), thus showing that the more time ants spent on the information transmission, the more complex – in the sense of Kolmogorov complexity – was the message. This surprisingly resembles "learning with chunking and generalization" during the foraging process in structured environments suggested by Kolodny et al. (2015). In sum, ideas from information theory help to evaluate cognitive and flexible aspects of natural communication systems and thus understand better the evolution of language.

## Acknowledgements

This work was funded by Russian Fund for Basic Research (No. 20-04-00072) and the Federal Fundamental Scientific Research Program for 2021-2025 (122011800268-1).

## References

Bickerton, D. (1990). *Language and species*. Chicago: The University of Chicago Press.

Fischer, J. (2020). Nonhuman primate alarm calls then and now. *Animal Behavior and Cognition*, *7*(2), 108-116.

Girard-Buttoz, C., Zaccarella, E., Bortolato, T., Friederici, A. D., Wittig, R. M., & Crockford, C. (2022). Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties. *Communications biology*, *5*(1), 1-15.

Gong, T., Shuai, L., & Comrie, B. (2014). Evolutionary linguistics: theory of language in an interdisciplinary space. *Language Sciences*, *41*, 243-253.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., ... & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in psychology*, *5*, 401.

King, S. L., & Janik, V. M. (2013). Bottlenose dolphins can use learned vocal labels to address each other. *Proceedings of the National Academy of Sciences*, *110*(32), 13216-1322

Kohl, P. L., Thulasi, N., Rutschmann, B., George, E. A., Steffan-Dewenter, I., & Brockmann, A. (2020). Adaptive evolution of honeybee dance dialects. *Proceedings of the Royal Society B*, *287*(1922), 20200190.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of information transmission*, *1*(1), 1-7.

Kolodny, O., Edelman, S., & Lotem, A. (2015). Evolution of protolinguistic abilities as a by-product of learning to forage in structured environments. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1811), 20150353.

Reznikova, Z., & Ryabko, B. (2011). Numerical competence in animals, with an insight from ants. *Behaviour*, *148*(4), 405-434.

Reznikova, Z. (2021). Ants' Personality and Its Dependence on Foraging Styles: Research Perspectives. *Frontiers in Ecology and Evolution*, *9*, 269.

Ryabko, B., Reznikova, Z. (2009). The use of ideas of information theory for studying "language" and intelligence in ants. *Entropy*, *11*(4), 836-853.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379-423.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Wacewicz, S., & Żywiczyński, P. (2015). Language evolution: Why Hockett’s design features are a non-starter. *Biosemiotics*, *8*(1), 29-46.

# INVESTIGATION OF OCTAVE SIMILARITY IN PITCH PERCEPTION IN RATS : IN SEARCH OF COMMON AUDITORY PROPERTIES BETWEEN HUMANS AND ANIMALS

RISERU KOSHIISHI*[1] and KAZUO OKANOYA[1,2]

*Corresponding Author: koshiishi-riseru5414@g.ecc.u-tokyo.ac.jp
[1]Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
[2]Teikyo University Advanced Comprehensive Research Organization, Tokyo, Japan

Pitch is the psychological representation of tone height. Pitch is a very important sound attribute that conveys information in speech, music, and animal vocalizations (Doupe & Kuhl,1999; Rothenberg et al., 2014). Various models have been proposed for how we perceive pitch, but they have yet to be fully demonstrated.

It has been found that mental representations of pitch are more likely to explain various phenomena if they are multidimensional. The spiral structure model (Shepard, 1964) is based on the theory that there are two aspects of pitch perception: pitch height, which has a linear relationship with fundamental frequency, and pitch chroma, which is a repeated sequence of notes. These aspects are illustrated in the model of a sound spiraling up the outside of a cylinder. In this model, one revolution of the pitch chroma represents an octave.

When two sounds have a frequency ratio of 1:2, i.e., are in octaves, the perceptual impression of the two sounds is similar. This phenomenon called octave similarity. Octave similarity forms the basis of the pitch chroma concept and can be said to be the foundation for examining pitch perception models. Octave similarity is known to be a general characteristic that is not limited to musical experience and has been observed in infants and in rhesus monkeys (Demany & Armand, 1984; Wright et al., 2000). This suggests that octave similarity is common in a wide range of animal species, not only humans, and that it may have a biological origin. However, the mechanism and evolutionary function of octave similarity are still unclear.

In this study, we used the rat (*Rattus norvegicus*), which is frequently used as a model animal for speech communication, to aim to investigate the mechanism and biological origin of octave similarity in more detail. Although the detection of octave similarity has been explored in other species other than primates, it has not yet been fully demonstrated (Deutsch, 1943; Burns, 1999; Cynx,1993). We conducted behavioral experiments using operant conditioning to elucidate whether octave similarity can be detected in the auditory system of rats.

The experiment was conducted in an operant apparatus. Five rats were first rewarded with a sucrose solution to form left and right lever-pressing behavior. Two groups of octave-related sound stimuli (A: 1250 Hz, 2500 Hz, 5000 Hz; B: 1000 Hz, 2000 Hz, 4000 Hz) were set up, and the rats were trained to discriminate between left and right lever presses according to the pitch of these sound stimuli. Finally, the rats were probed with a sound 1-2 octaves away from the sound presented in the training.

As a result of the discrimination training, all rats achieved the training criterion and were able to perform the probe test twice. As a result of probe tests, all rats perceived the probe sounds and showed left or right lever-pressing responses. However, the results did not show octave similarity, i.e., similarity of perception of the probe sound to that of the training sound, and the rats showed different responses depending on the individual and the probe sounds. The analysis of response latency suggested that compared to the sound stimuli presented in the training trials, probe sounds were perceived as different, and this tendency was particularly pronounced for probe sounds that were two octaves apart.

When we examined the difference in response shown by rats, it was suggested that comparing the relative pitch of the presented sound stimuli may have caused a deflection of the left/right response. Furthermore, it was considered that some rats responded not only to the presented sound stimuli but also to left/right preference and past response content in lever pressing during the session as a cue during the probe test. From now on, we will adjust the experimental method to eliminate response cues other than the presented sounds. In addition, we plan to conduct a test in which rats are trained to discriminate a part of the probe sounds and then presented with another new probe sounds, in order to further verify the phenomenon of pitch perception.

The progress of this study will allow us to examine the possibility that octave similarity is a common mechanism in many animal species, and will provide new insights into auditory research. Furthermore, the possibility that octave equivalence is related to the specificity of human speech signals can be examined.

## References

Cynx, J. (1993). Auditory frequency generalization and a failure to find octave generalization in a songbird, the European starling (Sturnus vulgaris). *Journal of Comparative Psychology*, *107*(2), 140.

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *The Journal of the Acoustical Society of America*, *76*(1), 57-66.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, *22*(1), 567-631.

Dowling, W. J., & Harwood, D. L. (1986). *Music Cognition*. New York: Academic Press.

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics*, *11*(6), 411-412.

He, C., & Trainor, L. J. (2009). Finding the pitch of the missing fundamental in infants. *Journal of Neuroscience*, *29*(24), 7718-8822.

Hoeschele, M. (2017). Animal pitch perception: melodies and harmonies. *Comparative Cognition & Behavior Reviews*, *12*, 5.

Richards, D. G., Wolz, J. P., & Herman, L. M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, Tursiops truncatus. *Journal of Comparative Psychology*, *98*(1), 10.

Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., & Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hearing Research*, *308*, 71-83.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, *36*(12), 2346-2353.

Titze, I. R., & Palaparthi, A. (2020). Vocal loudness variation with spectral slope. *Journal of Speech, Language, and Hearing Research*, *63*(1), 74-82.

Tomlinson, R. W., & Schwarz, D. W. (1988). Perception of the missing fundamental in nonhuman primates. *The Journal of the Acoustical Society of America*, *84*(2), 560-565.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General*, *129*(3), 291.

# THE COMMON TASK FRAMEWORK: USING CAUSAL THEORIES FROM LINGUISTIC ANTHROPOLOGY TO EXPLORE THE EMERGENCE OF SYMBOLIC COMMUNICATION

SEÁN G. ROBERTS*[1], ELIZABETH IRVINE[1], FIONA M. JORDAN[2]

*Corresponding Author: RobertsS55@cardiff.ac.uk
[1]English, Communication and Philosophy, Cardiff University, Cardiff, UK
[2]Anthropology and Archaeology, University of Bristol, Bristol, UK

We present the Common Task Framework approach to testing causal theories about the evolution of language. There are now many theories about how symbolic communication emerged, but less work trying to compare, synthesise and test these theories. We suggest that the first step is to formalize the theories as causal graphs using tools from the field of causal inference. This helps recognize the critical causal links that differentiate theories. The second step is to use tools from experimental semiotics to specify a "common task": an experimental environment and a task for individuals to complete. The different theories suggest different design solutions for this common task, and the success of the individuals can be used as a measure of the relative success of each theory. In this paper, we provide an example from anthropological theories of the emergence of symbolic communication, including coding 11 theories as causal graphs using the CHIELD database. We hope the common task framework will be applicable to many aspects of language evolution.

## 1. Introduction

Human language is the most complex communication system on earth, with no other animal coming close to our linguistic abilities. How did this ability evolve, and why only in humans? While genetic and cognitive differences are obviously part of the story, perspectives from anthropology and archaeology are making it increasingly clear that our ancestors also benefitted from critical social, economic and ecological situations. Some species including apes, birds and even insects have advanced communicative abilities (Berwick et al., 2011; Arnold & Zuberbühler, 2006; von Frisch, 1967) or can acquire them through intense training (e.g. Premack, 1971). However, apparently their natural habitats do not provide the right selective pressures for these latent abilities to evolve further. Only for humans did the right factors come together to motivate the evolution of complex language. Therefore, we argue that there is a critical unanswered question: *What* *are the social, economic and ecological conditions for the evolution of complex communication systems?* (Roberts, 2018). We suggest that theories from **anthropology and archaeology** can help us answer this question.

There are two main challenges. The first is identifying the most prominent theories. More importantly, how do they agree or conflict with each other? Are there specific predictions that can be tested in order to evaluate the relative plausibility of each theory? Addressing this will require a systematic review of the literature and a way of formally representing the causal structure of theories. We suggest that **causal inference** can help researchers to do this.

The second challenge is how to test these theories against each other empirically. Since we cannot observe language evolution directly, and naturalistic methods provide limited flexibility, support for theories must come from a robust combination of approaches (Irvine, Roberts & Kirby, 2013). We suggest that a **common task framework** is needed that combines control from experimental semiotics and ecological validity from anthropology. This should be flexible enough to simulate many scenarios while being consistent enough to make comparisons across theories.

In this paper we outline how causal inference and a common task framework can meet the two challenges above. We illustrate the approach with examples related to the emergence of symbolic signals. However, because many questions about the evolution of language involve complex causal connections from multiple sources of evidence, and where direct experimentation is impossible, we hope that the common task framework will be a useful more generally for the field of evolutionary linguistics.

## 2. Formalising theories from anthropology and archaeology

There is now a wealth of theories from anthropology and archaeology about the kinds of tasks that plausibly created a need for complex communication in human evolutionary history (see e.g. Dor et al., 2014; Power et al., 2016). These include predation (Dunbar, 2017), hunting (Knight & Lewis, 2017; Sterelny, 2012), navigation (Bednarik, 1997), tool use (Davidson & Noble, 1989), fire making (Twomey, 2013; Wiesser, 2014), and reproductive strategies (Knight et al., 1995). While theories abound, there are few attempts to synthesise them and systematically test them against each other. We argue that causal inference provides a powerful approach that can express, explore and evaluate theories. One tool for helping researchers do this is the Causal Hypotheses in Evolutionary Linguistics Database (CHIELD, Roberts et al., 2020, http://chield.excd.org). CHIELD is a database of hypotheses which have been hand-coded as causal

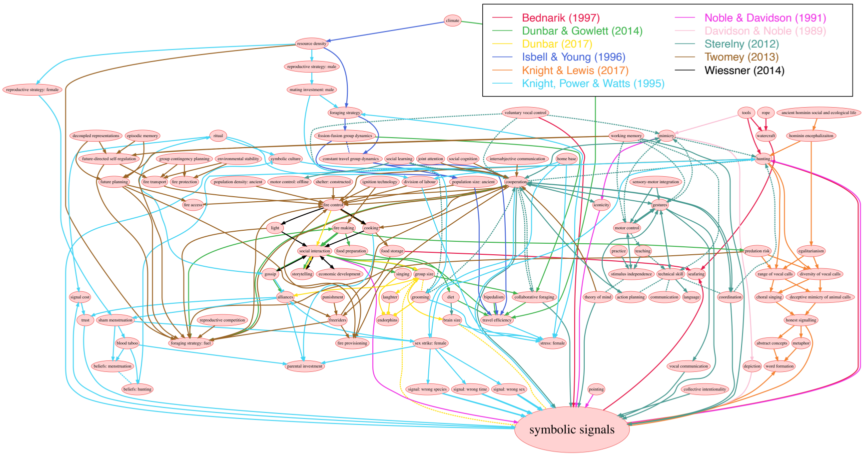


*Figure 1: A causal graph of 11 theories of the emergence of symbolic signals.*

graphs. This allows users to discover formal links between hypotheses, assess conflicting evidence and spot weak links that currently have little support.

Using CHIELD, we formalized 11 theories as causal graphs (Fig. 1, see supporting materials). There were two major findings from this exercise. The first is that the theories are deeply interconnected. For example, the move from dense jungle out into open savannah could have changed the availability of resources and created pressures to change foraging strategies and group dynamics (Isbell & Young, 1996). This could have made larger social networks more feasible, and created greater pressures for advanced social communication (Dunbar, 2004). However, a complex infrastructure that facilitates trust and material support within a group may have been necessary before symbolic communication was viable (Knight et al., 1995). These infrastructures themselves may have been shaped by specific ecological and economic conditions, such as the ability to harness the environment more effectively by crafting tools (Gibson et al., 1994; Kolodny et al., 2018), collaborative hunting (Sterelny, 2012), or the need to maintain sources of fire (Twomey, 2013). We also noted some links that, as far as we know, have not been remarked on in the literature: Knight et al.'s ritual theory involves beliefs about hunting; hunting animals provides bones that can be used as fuel for fires; fire is used to reduce predation risk.

The second finding is that, despite the interconnectedness, the majority of theories are not mutually exclusive. There are only two points of formal conflict in the causal graph. Firstly, Knight et al. do not think that gossip on its own can create effective alliances, in contrast to Dunbar. Secondly, Noble & Davidson suggest that symbolic language facilitates hunting, rather than the other way around. Finally, around half of the causal connections were hypothesized in the source papers without empirical support. This suggests there is much outstanding work to make these theories robust. In order to accomplish this work, we need a practical testing method, for which Experimental Semiotics is a candidate.

## 3. Experimental semiotics

Experimental semiotics uses methods from psychology to explore the role of cognition, acquisition, and usage in shaping language (Galantucci, 2009; Roberts, 2017). Studies use lab-based experiments with human participants who must construct, use, and transmit artificial languages. These methods have been used to explore the emergence of compositional structure (Kirby et al., 2015). Initial results suggested that inter-generational transmission was a key causal property, but recent studies test whether the same pressure can be created by needing to communicate with a large population (Raviv, Meyer & Lev-Ari, 2019). Essentially, the experiments provide a way of comparing design solutions to the "common task" of creating conditions for the emergence of compositionality.

However, in order to obtain a high level of experimental control, these experiments tend to use very idealised tasks based on formal communication games, and impose strict limits on the ways that individuals can interact: one individual speaks at a time; pointing is only allowed during feedback; or individuals can't initiate repair (see Macuch Silva & Roberts, 2016). Irvine & Roberts (2016) noted that several experimental semiotics studies assume that the pressure to communicate derived from the need for individuals to collaborate in order to manipulate physical objects or build complex constructions. However, the experimental tasks are so constrained that they remove the possibility of any other strategies. We argue that a different kind of common task is needed to explore the origins of symbolic communication.

## 4. A common task framework for studying symbolic communication

A common task framework approach involves iterating on the solution of a practical task in order to understand how different pressures lead to different outcomes. For example, in the field of robotics, the *DARPA Grand Challenge* is a common task framework where an autonomous robotic vehicle must complete a task in the real world such as driving across a desert. Different robotic designs can be evaluated against each other according to how well they complete the task. While the DARPA Challenge focuses on designing robots, for studying the emergence of symbolic communication we suggest that the experimenter's task is to design a **situation** (including a physical **environment** and **survival task** that individuals must complete) that creates a pressure for individuals to develop a symbolic communication system. The effectiveness of the situation is tested by placing agents (with a capacity for symbolic communication, but no existing conventions within the challenge context) into the situation and observing if they do indeed invent a symbolic communication system.

These situations may be somewhat abstract, but they should reflect relevant and plausible analogues of early hominid life. Using the causal graphs as a guide, different theories are translated into minimally contrasting situations. These situations are then tested (using human participants or simulated agents), and the results can be contrasted to evaluate the theories against each other. The aim of the common task framework is not to recreate the exact, true emergence of symbolic communication in human ancestors: we cannot know what this situation was. Instead the procedures are used as grounded thought experiments in order to explore and refine theories (e.g. Webb, 2000; Steels, 2003).

One candidate framework for studying the emergence of symbolic communication comes from Irvine & Roberts (2016). This was an experiment with two human participants in a 3D virtual world, using the video game *Minecraft*. The environment was a field outdoors, and the survival task was to build a shelter together. Each participant had half of the plan for the shelter, providing a reason to communicate. Participants were prevented from using speech during the experiment, but they were given the resources and capacity to construct a simple symbolic communication system to help them in the task: they could use their virtual avatar to point and they could also knock on the table. This gave them the option of developing a simple symbolic system to refer to the four different building materials (e.g. one knock for red blocks, two knocks for yellow blocks etc.). That is, participants were placed in a similar situation to our pre-linguistic ancestors: they had the capacity to construct a symbol system, but the crucial question was whether the task would motivate them to do so. The results showed that all participants used pointing and trial-and-error strategies to successfully complete the task. Post-experiment interviews confirmed that there was not enough motivation to invent a symbolic system, and so none evolved. After all, setting up a symbolic system takes time, and the pointing strategy was productive enough to be successful. Irvine & Roberts concluded that the need to collaborate for construction would not have been a strong enough selective pressure to motivate the emergence of a symbolic system for our ancestors.

In summary, a pressure to develop a complex communication system relies on a particular practical situation. It might seem obvious that various theories would create a pressure for the emergence of symbolic communication. However, by putting the theory into practice, Irvine & Roberts realized that the opportunity cost of setting up a symbolic system is high, and participants may find unexpected solutions that avoid having to do so. The question now is: what situations would motivate them to invest time in creating a symbolic system?

## 5. Candidate situations

Based on the causal review of the literature, we propose that situations which promote the emergence of symbolic communication will have various key properties. First, there will be an **asymmetry of information** between individuals. That is, some individuals will know facts or skills that the others do not, creating something to communicate about. A second property is that the referents to be communicated about will be **distant in time or space**, making pointing less effective. These first two properties are found in some advanced non-human communication systems, such as honey bees needing to communicate the location of pollen (von Frisch, 1967). Another key property is the need for **division of labour**. This requires coordination and scheduling between individuals, requiring trust and cooperation. Principles from video game design (Brown, 2018) suggest more key properties, including: **More tasks than people** to promote role-switching; **dead time** to promote multitasking; **disruptions** which change task demands to prevent ritualization of strategies; **public goods** to promote coordination of who does tasks with no direct reward. We predict that these key properties will emerge as common elements in successful situation designs. These will define a set of properties that solve the "meta-task" of promoting the emergence of symbolic communication.

There are two candidate situations motivated by anthropological theories that may meet these key properties. The first involves fire maintenance. Humans began using fire long before being able to create it (Twomey, 2013), and even today fire *making* technology is not universal in human societies (McCauley, Collard & Sandgathe, 2020). Keeping a fire lit and fuelled create strong pressures to organise and divide labour between individuals. This could provide the right pressure to start referring to distant locations (sources of fuel) or points in time (agreements about tending the fire). The situation also involves dead time, disruptions and public goods. Manipulations of this situation could include how far away and how sparsely distributed the fuel sources were. The most successful experimental conditions could then be compared to plausible scenarios of early hominids.

The second situation is collaborative hunting. Sterelny (2012) suggests that the potential rewards of hunting big game would provide pressures to cooperate and communicate. An asymmetry of information exists between skilled and unskilled hunters. Some hunting strategies involve different roles (e.g. flushing and ambushing), and these roles may require individuals to be out of sight of each other at key moments. This would create a need for division of labour and referring to future times and distant places in order to plan the hunt. Conditions could vary the number of participants or the speed of the animal. Fast animals might encourage strategies involving stealth or splitting the group into different roles, requiring communication about locations, time and coordinated action. The core causal components of these candidate situations can be recreated in a common task framework. We note that *Minecraft* has various features that help with this: fire dynamics; hearths that require a constant supply of fuel; projectile weapons; and animals that can be 'hunted'. It also supports computational agents.

## 6. Discussion

We presented a common task framework approach to investigating the evolution of symbolic communication. Causal inference tools were used to formally relate theories and identify critical differences. These were used to design a common task framework to evaluate those theories against each other. We suggested that theories related to fire maintenance and collaborative hunting meet some key properties for the emergence of symbolic communication. Of course, insights from this approach need to be integrated with other findings, including the role of repeated actions leading to conventionalisation and differences in cognitive abilities between our ancestors and modern human participants.

## Acknowledgements

Supported by AHRC grant AH/T006927/1. We thank our reviewers.

## References

Arnold, K. and Zuberbühler, K. (2006). Language evolution: semantic combinations in primate calls. Nature, 441(7091):303–303.

Bednarik, R. G. (1997). The origins of navigation and language. Artefact, 20(1997):16.

Berwick, R. C., Okanoya, K., Beckers, G. J., and Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. Trends in Cog. Sci., 15(3):113–121.

Brown, M. (2018) How Overcooked's Kitchens Force You to Communicate. https://www.youtube.com/watch?v=C3M8BvWcJQY

Burkart, J. M., Hrdy, S. B., and Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. Evolutionary Anthropology, 18(5):175–186.

Dor, D., Knight, C., Lewis, J. (2014) The social origins of language. OUP.

Dunbar, R. I. (2004). Gossip in evolutionary perspective. R. Gen. Psy. 8(2):100.

von Frisch K (1967). The Dance Language and Orientation of Bees. HUP.

Galantucci, B. (2009). Experimental semiotics: A new approach for studying communication as a form of joint action. Topics in Cog. Sci., 1(2), 393-410.

Gibson, K.R. & Ingold, T. (1994). Tools, language and cognition in human evolution. CUP.

Hill, R. A., & Dunbar, R. I. (1998). An evaluation of the roles of predation rate and predation risk as selective pressures on primate grouping behaviour. Behaviour, 135(4), 411-430.

Irvine E. and Roberts S. G. (2016). Deictic Tools Can Limit The Emergence Of Referential Symbol Systems. In S.G. Roberts, et al. (eds.) The Evolution of Language: Proceedings of the 11th International Conference.

Irvine, L., Roberts, S., & Kirby, S. (2013). A robustness approach to theory building. In Proc. Cognitive Science Society (Vol. 35, No. 35).

Isbell, L. A. and Young, T. (1996). The evolution of bipedalism in hominids and reduced group size in chimpanzees. J. Human Evolution, 30:389–397(9)

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. Cognition, 141, 87-102.

Knight, C. and Lewis, J. (2017). Wild voices: Mimicry, reversal, metaphor, and the emergence of language. Current Anthropology, 58(4).

Knight, C., Power, C., & Watts, I. (1995). The human symbolic revolution: a Darwinian account. Cambridge archaeological journal, 5(1), 75-114.

Kolodny, O. and Edelman, S. (2018). The evolution of the capacity for language. Phil. Trans. R. Soc. B, 373(1743):20170052.

Macuch Silva V. & Roberts S. (2016) Exploring the Role of Interaction in the Emergence of Linguistic Structure. In S. Roberts & G. Mills (Eds.) Proceedings of EvoLang XI, Language Adapts to Interaction Workshop.

Macuch Silva V. & Roberts S. (2016) Exploring the Role of Interaction in the Emergence of Linguistic Structure. In S. Roberts & G. Mills (Eds.) Proceedings of EvoLang XI, Language Adapts to Interaction Workshop.

McCauley, B., Collard, M., & Sandgathe, D. (2020). A cross-cultural survey of on-site fire use by recent hunter-gatherers. J. Paleolithic Arch., 3(4), 566-584.

Power, C., Finnegan, M., Callan, H. (2016) Human origins: contributions from social anthropology. Berghahn Books.

Premack, D. (1971). Language in chimpanzee. Science, 172(3985), 808-822.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. Proc. Royal Society B, 286(1907), 20191262.

Roberts, G. (2017). The linguist's Drosophila: Experiments in language change. Linguistics Vanguard, 3(1).

Roberts, S. (2018). What are the social, economic and ecological conditions for the evolution of complex communication systems? Phys L. Rev., 26,152-154.

Roberts, S. G., Killin, A., Deb, A. et al. (2020). CHIELD: the causal hypotheses in evolutionary linguistics database. J. Language Evolution, 5(2), 101-120.

Steels, L. (2003). Evolving grounded communication for robots. Trends in cognitive sciences, 7(7), 308-312.

Sterelny, K. (2012). Language, Gesture, Skill: The Co-Evolutionary Foundations of Language. Phil. Trans. R. Soc. B, 367(1599), 2141–2151.

Twomey, T. (2013). The cognitive implications of controlled fire use by early humans. Cambridge Archaeological Journal, 23(1):113–128.

Webb, B. (2000). What does robotics offer animal behaviour? Animal behaviour, 60(5):545– 558.

# THE EVOLUTIONARY TRENDS OF NOUN CLASS SYSTEMS IN ATLANTIC LANGUAGES

Neige Rochant[1], Marc Allassonnière-Tang[2], and Chundra Cathcart[*3]

[*]Corresponding Author: chundra.cathcart@uzh.ch
[1]Sorbonne Nouvelle University/Lacito, CNRS, Paris, France
[2]EA, MNHN/ CNRS/ University Paris City, Paris, France
[3]Department of Comparative Language Science, University of Zurich, Zurich, Switzerland

Nominal classification systems such as grammatical gender (e.g., the masculine/feminine distinction in French) and noun classes (e.g., Bantu noun classes based on fruits, plants, liquids, among others) provide a window on how the human brain perceives and categorizes objects and experiences it encounters. While the diachronic development of grammatical gender systems is well studied, noun class systems have received less attention. We use phylogenetic comparative methods to analyze where noun classes are marked (on nouns, pronouns, demonstratives, articles, adjectives, numbers, and verbs) in thirty-six Atlantic languages and how these markers change diachronically. Our results show that noun class marking is generally preferred and more stable within the noun phrase, i.e., on nouns, demonstratives, and adjectives.

## 1. Introduction

Languages can rely on various strategies to categorize nouns of the lexicon (Seifart, 2010; Kemmerer, 2017). One the most common strategies is noun class systems (Corbett, 2007), in which each noun of the lexicon is assigned to a specific category (i.e., noun class), which can relate to humans, plants, fruits, liquids, among others (Corbett, 2013). For example, in Swahili, nouns are affiliated to more than ten noun classes.

Two of the most common formal criteria to define noun classes are flexibility of category assignment and grammatical agreement (Corbett, 1991). First, noun class systems typically have a rigid assignment system. Each noun of the language belongs to one of the noun classes found in the language. This assignment is not flexible and using the agreement pattern of another noun class would result in ungrammaticality. Second, noun class systems may have marking on nouns, as shown in Swahili, with prefixes on the nouns, e.g., *m-toto* (CLASS.1-child) ‘child’. However, noun classes also generate grammatical agreement with words associated to the noun and its referent. The noun classes can be marked on the adjectives, verbs, demonstratives, numerals, among others. Taking Swahili again to illustrate this type of agreement: *m-toto yu-le a-li-anguka* (CLASS.1-child CLASS.1-that CLASS.1.SUBJ-past-fall) ‘that child fell’. These requirements of assignment and agreement distinguishes noun class systems from inflection classes such as on number marking. Grammatical number systems have agreement; however, number is not a fixed inherent feature of a noun/referent. For example, count nouns can generally be singular or plural depending on the context while the noun class of a noun is fixed. It is generally held that languages may develop a nominal classification system that proceeds through different stages of grammaticalization (Grinevald, 2002; Aikhenvald, 2016). At the beginning, classification is based on lexical nouns, which then develop into classificatory morphemes, which in turn further grammaticalize and become agreement markers. For example, the Niger-Congo noun class systems may have developed from nouns, along with the nominal classification systems found in modern Amazonian languages (Grinevald & Seifart, 2004). At the same time, questions remain regarding their development that are relevant to the diachronic change of linguistic complexity (Walchli, Olsson, & Di Garbo, 2020). For example, were noun classes first marked on nouns? If so, how did they spread to pronouns or verbs? While several studies have investigated the synchronic distribution of noun class systems in languages of the world (Corbett, 2013; Allassonnière-Tang et al., 2021), most diachronic quantitative studies have focused on the evolution of agreement marking in Indo-European languages (Allassonnière-Tang & Dunn, 2020; Carling & Cathcart, 2021). Few studies have investigated the evolution of agreement marking in languages with noun class systems, and even fewer studies have approached this question from a quantitative perspective with phylogenetic methods. The current study aims to fill this gap, quantifying the diachronic preference, speed of change, and stability of noun class markers across multiple morphosyntactic domains in the Atlantic languages of West Africa, with an eye to inferring the relative chronology of their development.

## 2. Data

In this section, we describe how the data on noun class marking was gathered in a sample of 36 Atlantic languages. We also provide information about how the phylogeny of the languages in the sample was generated.

### 2.1. *Noun class systems*

The Atlantic languages were selected due to their frequent presence of noun class systems. We consider noun class marking on the noun, in the noun phrase, and the verb phrase. To be more precise, we consider noun class marking on the noun itself (via a PREFIX), ARTICLES, PRONOUNS, DEMONSTRATIVES, ADJECTIVES, CARDINAL NUMBERS, and VERBS. For instance, Segerer (2002, 85-92) shows that Bijogo (Glottocode `bidy1244`) has fourteen noun classes that are marked on the noun itself (*e-we* [class.E-goat] ‘goat’), adjectives and demonstratives (*ɲo-ogo n-nɛ n:-gbon* [CLASS.M-rock CLASS.M-DEM CLASS.M-be.big] ‘These rocks are big.’), pronouns (*ya-g* [CLASS.YA-PR] ‘them’), cardinal numbers (*ya-to ya-nsom*

[CLASS.YA-two CLASS.YA-people] ‘two people’), and verbs (*bisaw wɔ-gbe na-jɔkɔ na-kotong* [Bissau CLASS.WO-ACC.have CLASS.ÑA-house CLASS.ÑA-big] ‘There are big houses in Bissau.’). A language is encoded as marking a domain as long as one instance of marking is attested. Data were extracted manually from language grammars and sketches with sufficient available information, while maintaining a balance across the sub-branches of the language family. This process resulted in a sample of 36 languages, displayed in Figure 1.

### 2.2. *Tree Sample*

The models used in this paper require a phylogenetic representation of the languages in our sample, in the form of a tree sample. We matched each speech variety in our data set to its closest correspondent in a data set of automatically generated lexical cognacy and sound class characters (Jäger, 2018). We inferred a phylogeny of the Atlantic languages using RevBayes (Höhna et al., 2016), using a Birth-Death tree prior (Yang & Rannala, 1997) and a General Time-Reversible model of character evolution (Tavaré, 1986). We employed clade constraints, enforcing a split between the North and Bak languages and including the seven higher-order subgroups found in Glottolog (Cangin, Central Atlantic, Fula-Sereer, Jaad, Naluic, Tenda, and Wolof-BKK), ensuring that trees that do not contain these subgroups are assigned zero posterior probability. We run 500,000 iterations of Markov chain Monte Carlo over 4 chains, discarding the first half of samples as burn-in and monitoring convergence by comparing the log posteriors of the chains.

## 3. Method

We explore differences in the diachronic behavior of noun class marking across the seven domains of marking in our sample via phylogenetic comparative methods. We assume that marking in each domain evolves independently according to a continuous-time Markov process parameterized by two rates, a gain rate $\alpha_d$ and loss rate $\beta_d : d \in \{1, ..., 7\}$. We infer the rates for all features jointly using RStan (Carpenter et al., 2017). We place $\text{Gamma}(\lambda, \lambda)$ and $\text{Gamma}(\mu, \mu)$ priors over $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, where $\lambda, \mu \sim \text{Exp}(1)$, and incorporate phylogenetic uncertainty by inferring rates for 100 trees from our tree sample and aggregating posterior samples of rates across trees. We use posterior rate values in order to generate a number of quantities of interest to the properties mentioned above, as described below. Code employed in this paper is available at `https://github.com/chundrac/JCoLE2022-atlantic`.

**Stationary probabilities of noun class marking** We make use of the stationary probability of noun class marking in each marking domain in order to operationalize the LONG-TERM PREFERENCE FOR NOUN CLASS MARKING across different morphosyntactic elements. The stationary probability of a continuous-time Markov chain represents the probability that the system will be in a particular

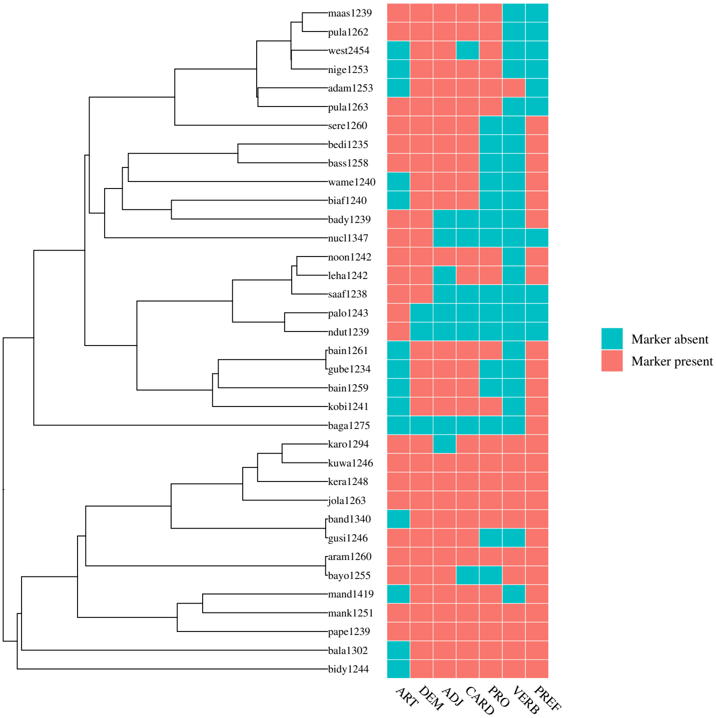


Figure 1. Maximum clade credibility tree of languages in data set, along with data

state as time approaches infinity, as well as the long-term preference of the system for a particular state. For a binary feature with a gain rate $\alpha$ and loss rate $\beta$, the stationary probability is equal to $\frac{\alpha}{\alpha+\beta}$. We compute posterior stationary probabilities of noun class marking in each domain from the posterior distribution of rates.

**Speed of change** Inferring gain and loss rates also allows us to compute the OVERALL SPEED OF CHANGE, irrespective of the direction of change for all posterior samples. Given a gain rate $\alpha$ and a loss rate $\beta$, the overall speed of change is $\alpha + \beta$.

**Phylogenetic stability** We carry out ancestral state reconstruction for the internal nodes of the tree on the basis of the inferred rates. For each pair of rates in the posterior sample, we compute the probability of noun class marking at each node in the tree and draw a Bernoulli variate indicating the presence or absence of noun class marking. We then average these values, yielding a posterior probability of presence between 0 and 1. On the basis of these reconstruction probabilities, we infer the phylogenetic stability of marking in each domain according to the Bayesian method of Borges et al. (2019), which provides an index of a feature's relative invariance over time, regardless of whether it is preferred or dispreferred. We infer posterior values of the stability metric jointly across marking domains for all trees in the sample.

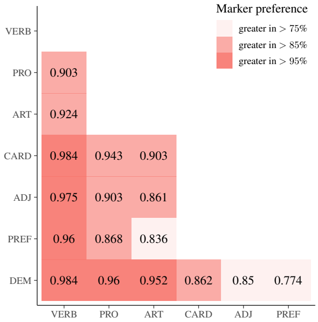


Figure 2. Cross-sample pairwise differences in long-term preference for noun class marking between marking domains

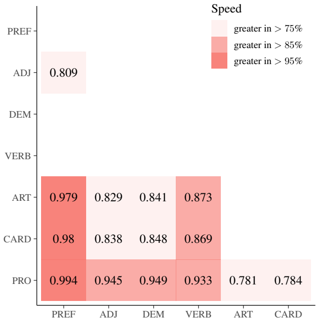


Figure 3. Cross-sample pairwise differences in speed of change between marking domains

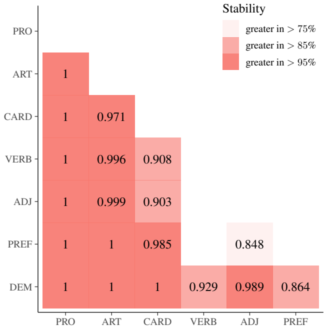


Figure 4. Cross-sample pairwise differences in stability between marking domains, across samples

## 4. Results

We compare posterior distributions or each metric of interest across pairs of domains. Following Gelman, Hill, and Yajima (2012), feature pairs display decisively contrasting behavior if differences between values in 95% or more of samples are greater/less than zero.[1] The 85% cutoff serves as strong but not decisive evidence. Inter-feature comparisons for long-term noun class marker preference, speed, and stability can be found in Figures 2, 3 and 4, respectively. Features are organized according to their median posterior values, in increasing order.

The following cline can be found with respect to long-term noun class marking: VERB < PRO, ART < CARD, ADJ, PREF, DEM (where << indicates decisive and < indicates strong evidence for a difference). There is additionally decisive evidence that cardinal numbers, adjectives, prefixes and demonstratives show greater preference for noun class marking than verbs and that demonstratives show greater preference for noun class marking than pronouns and articles.

Prefixes exhibit the lowest speed of change, followed by adjectives, demonstratives, verbs, articles, cardinal numbers and pronouns. Unlike preference for noun class marking, no decisive breaks can be found along this cline. There is decisive evidence that articles, cardinal numbers, and pronouns exhibit faster change than prefixes; additionally, there is strong evidence that pronouns exhibit faster change than adjectives, demonstratives, and verbs, and that articles, cardinal numbers and pronouns exhibit faster change than verbs.

In some ways, results for stability mirror those of speed, with pronouns, articles and cardinal numbers exhibiting lowest stability and highest speed, verbs exhibiting intermediate values for both metrics, and adjectives, prefixes and demonstratives exhibiting highest stability and lowest speed. Here, however, we see decisive and strong evidence for breaks along the cline of stability: PRO << ART << CARD, VERB, ADJ < PREF, DEM.

## 5. Concluding discussion

The high preference and stability for demonstratives is in line with the literature. From a diachronic point of view, demonstratives are frequently the source of grammaticalization processes in Atlantic languages. Demonstratives inflected for class frequently grammaticalize as class pronouns, which then reinforce and extend class agreement (Creissels & Pozdniakov, 2015). The high preference and stability for marking on nouns also dovetails with received wisdom. For example, in noun class languages that lost grammatical agreement, markers are generally still found on nouns (Kießling, 2018). The lower preference for noun class marking on verbs is expected as noun class marking typically starts from within the noun phrase and then extends to the verb phrase (Tang & Her, 2019). Results show that the marking on pronouns is the least stable and most in flux, which reflects that

---

[1] Debate exists regarding the danger of false discoveries under multiple comparisons in a Bayesian framework. Our hierarchical model has a partial pooling index of 0.67 (Ogle et al., 2019), which corresponds to a low rate of false positives but a high rate of false negatives. A full appraisal of this issue is outside of this paper's scope but will be addressed in future work, e.g., via a mixture model.

the presence of class pronouns is likely to interact with the grammaticalization of demonstratives and other diachronic factors.

The results also indicate that marking on demonstratives, prefixes, and adjectives is the most stable and preferred in comparison to articles, cardinal numbers, pronouns, and verbs. This speaks to the development of noun class marking first in nouns and demonstratives, subsequently spreading to other elements in the noun phrase and finally to the verbs. However, our models do not quantify the complex diachronic interactions between these different markers. For example, are noun class markers most likely to be found on nouns before demonstratives, or vice-versa? In future work, models of correlated evolution can be employed to further assess the interactive dynamics between different noun class markers.

## Acknowledgments

The first and second authors acknowledge the support of Sorbonne Nouvelle University and Lacito – UMR 7107, CNRS, and the French National Research Agency grant EVOGRAM (ANR-20-CE27-0021), respectively.

## References

Aikhenvald, A. Y. (2016). *How gender shapes the world.* Oxford: Oxford University Press.

Allassonnière-Tang, M., & Dunn, M. (2020). The evolutionary trends of grammatical gender in Indo-Aryan languages. *Language Dynamics and Change*, *Advance articles*, 1–30.

Allassonnière-Tang, M., Lundgren, O., Robbers, M., Cronhamn, S., Larsson, F., Her, O.-S., Hammarström, H., & Carling, G. (2021). Expansion by migration and diffusion by contact is a source to the global diversity of linguistic nominal categorization systems. *Humanities and Social Sciences Communications*, *8*(1), 331.

Borges, R., Machado, J. P., Gomes, C., Rocha, A. P., & Antunes, A. (2019). Measuring phylogenetic signal between categorical traits and phylogenies. *Bioinformatics*, *35*(11), 1862–1869.

Carling, G., & Cathcart, C. (2021). Reconstructing the evolution of Indo-European grammar. *Language*, *97*(3), 561–598.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

Corbett, G. G. (1991). *Gender.* Cambridge: Cambridge University Press.

Corbett, G. G. (2007). Gender and noun classes. In T. Shopen (Ed.), *Language typology and syntactic description Volume 3: Grammatical Categories and the Lexicon* (pp. 241–279). Cambridge: Cambridge University Press.

Corbett, G. G. (2013). Number of Genders. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Creissels, D., & Pozdniakov, K. I. (Eds.). (2015). *Les classes nominales dans les langues atlantiques.* Köln: Rüdiger Köppe Verlag.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of research on educational effectiveness*, *5*(2), 189–211.

Grinevald, C. (2002). Making sense of nominal classification systems: Noun classifiers and the grammaticalization variable. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 259–275). Amsterdam: John Benjamins.

Grinevald, C., & Seifart, F. (2004). Noun classes in African and Amazonian languages: Towards a comparison. *Linguistic Typology*, *8*(2), 243–285.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., & Ronquist, F. (2016). Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic biology*, *65*(4), 726–736.

Jäger, G. (2018). Global-scale phylogenetic linguistic inference from lexical resources. *Scientific Reports*, *5*.

Kemmerer, D. (2017). Categories of object concepts across languages and brains: the relevance of nominal classification systems to cognitive neuroscience. *Language, Cognition and Neuroscience*, *32*(4), 401–424.

Kießling, R. (2018). Niger-Congo numeral classifiers in a diachronic perspective. In W. B. McGregor & S. Wichmann (Eds.), *Diachrony of classification systems.* Amsterdam: John Benjamins.

Ogle, K., Peltier, D., Fell, M., Guo, J., Kropp, H., & Barber, J. (2019). Should we be concerned about multiple comparisons in hierarchical bayesian models? *Methods in Ecology and Evolution*, *10*(4), 553–564.

Segerer, G. (2002). *La langue bijogo de bubaque.* Paris: Peeters.

Seifart, F. (2010). Nominal Classification. *Language and Linguistics Compass*, *4*(8), 719–736.

Tang, M., & Her, O.-S. (2019). Insights on the Greenberg-Sanches-Slobin generalization: Quantitative typological data on classifiers and plural markers. *Folia Linguistica*, *53*(2), 297–331.

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of dna sequences. *Lectures on mathematics in the life sciences*, *17*(2), 57–86.

Walchli, B., Olsson, B., & Di Garbo, F. (2020). *Grammatical gender and linguistic complexity, Volume 1: General issues and specific studies.* Language Science Press. (OCLC: 1229599495)

Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using dna sequences: a markov chain monte carlo method. *Molecular biology and evolution*, *14*(7), 717–724.

# THE LANDSCAPE OF GESTURAL RESEARCH IN LANGUAGE EVOLUTION: A SYSTEMATIC REVIEW

Evelina D. Rodrigues[*1], António J. Santos[1], Flávia Veppo[2], Joana Pereira[3], and Catherine Hobaiter [4]

[*]Corresponding Author: erodrigues@ispa.pt
[1] William James Center for Research, ISPA–Instituto Universitário, Lisbon, Portugal
[2] Department of Applied Psychology, School of Psychology, University of Minho, Braga, Portugal
[3] Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[4] School of Psychology and Neuroscience, University of St Andrews, St Andrews, Scotland, UK

## 1. Introduction

The presence of common features in human and nonhuman primate communication can be used to suggest the evolutionary trajectories of potential precursors to language (Byrne et al., 2017; Fitch, 2010; Tomasello & Call, 1997). However, the overrepresentation of some species or populations[1] and differences in definitions, methodology and context distort our ability to make phylogenetic comparison (Fröhlich & Hobaiter, 2018; Leavens et al., 2019; Scott & Pika, 2012). This systematic review describes the current landscape of data available from studies of spontaneous gestural communication (produced without explicit training[2]) in human and nonhuman primates that make an explicit connection to language evolution.

## 2. Methods

---

[1] For example WEIRD - Western, Educated, Industrialized, Rich, and Democratic – in humans (Henrich et al., 2010); and BIZARRE - Barren Institutional Zoo And other Rare Rearing Environment – in chimpanzees (Leavens et al., 2010).

[2] Studies on home-sign development were included because of its often spontaneous emergence, but we excluded formal signed language studies to avoid explicit teaching of signs and gestures

We investigated how these studies varied in terms of study domains (Table S1), concept of gesture (Bourjade et al., 2020), study scope, inclusion of additional sources, and of research settings and study design (Table S2). In March 2020, we conducted a search of peer-reviewed articles and book chapters in Web of Science and PsycINFO using the search terms "gestur*," AND "evolutio*" OR "origin*," AND "languag*" OR "communicat*. From the 963 studies identified we retained 163 according to a predefined set of criteria (Figure S1).

## 3. Results

We found a similar number of studies on human (N=80) and nonhuman primates (N=87), but very few studies included data on both human and nonhuman primates (N=4). As a result, evolutionary inferences remain restricted to comparison across studies. We identified areas of focus, bias, and apparent gaps within the field. Of the nonhuman primate species studied, the majority focused on great apes (N=75/87), but no studies were found in small apes, monkeys of the Americas, or strepsirrhines (Figure S2). Most human participants were from WEIRD societies (N=68/80). There were few nonhuman primate studies of ontogeny and relatively few human studies of gesture form (Figure S3). We found variation in the conceptual and methodological approaches used between human and nonhuman primate studies. Definitions of gesture and criteria for intentional use are absent in most human studies (TableS3). Human studies focused more on specific gesture types or contexts (N=65/80) and were more likely to include additional sources to gesture (N=53/80). Studies of nonhuman primates were conducted more often in familiar settings (N=71/87) and using observational designs (N=61/87), whereas studies with humans were conducted more often in laboratories (N=53/80) using experimental designs (N=58/80; Table S4).

## 4. Discussion

Diversity in focus, methods, and socio-ecological context fill important gaps and provide nuanced understanding, but only where the source of any difference between studies is transparent. We highlight important areas in a call to action through which we can strengthen our ability to investigate gestural communication's contribution within the evolutionary roots of human language, including the need for: i) explicitly testing evolutionary hypotheses in our empirical work; ii) more data from diverse species, social groups, and environments, iii) studies exploring the spontaneous use of gesture forms in humans and gesture ontogeny in non-human primates.

## Acknowledgements

EDR was supported by the Portuguese Foundation for Science and Technology (SFRH/BD/138406/2018). CH is funded by the European Union's 8th Framework Programme, Horizon 2020, under grant agreement no 802719.

## References

Bourjade, M., Cochet, H., Molesti, S., & Guidetti, M. (2020). Is Conceptual Diversity an Advantage for Scientific Inquiry? A Case Study on the Concept of 'Gesture' in Comparative Psychology. *Integrative Psychological and Behavioral Science*, *54*(4), 805–832. https://doi.org/10.1007/s12124-020-09516-5

Byrne, R. W., Cartmill, E. A., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. E. (2017). Great ape gestures: Intentional communication with a rich set of innate signals. *Animal Cognition*, *20*(4), 755–769. https://doi.org/10.1007/s10071-017-1096-4

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Fröhlich, M., & Hobaiter, C. (2018). The development of gestural communication in great apes. *Behavioral Ecology and Sociobiology*, *72*(12), 194. https://doi.org/10.1007/s00265-018-2619-y

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Leavens, D. A., Bard, K. A., & Hopkins, W. D. (2010). BIZARRE chimpanzees do not represent "the chimpanzee." *Behavioral and Brain Sciences*, *33*(2–3), 100–101. https://doi.org/10.1017/S0140525X10000166

Leavens, D. A., Bard, K. A., & Hopkins, W. D. (2019). The mismeasure of ape social cognition. *Animal Cognition*, *22*(4), 487–504. https://doi.org/10.1007/s10071-017-1119-1

Scott, N. M., & Pika, S. (2012). A call for conformity: Gesture studies in human and non-human primates. *Developments in Primate Gesture Research*, 147–163.

Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford University Press, USA.

# COGNITION AND THE STABILITY OF EVOLVING COMPLEX MORPHOLOGY: AN AGENT-BASED MODEL

Erich Round[*1,2], Stephen Mann[1,3], Sacha Beniamine[1], Emily Lindsay-Smith[1], Louise Esher[4], and Matt Spike[5]

[*]Corresponding Author: e.round@surrey.ac.uk
[1]Surrey Morphology Group, University of Surrey, Guildford, UK
[2]School of Languages and Cultures, University of Queensland, St Lucia, Australia
[3]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[4]CNRS LLACAN, Paris, France
[5]School of Philosophy Psychology & Language Sciences, University of Edinburgh, Edinburgh, UK

Cultural attractors enable evolving cultural traits to gain the stability that underpins cumulative cultural evolution, yet the conditions that support their existence are poorly understood. We examine conditions affecting the stability of a salient kind of complex cultural attractor in human language, known as inflectional classes. We present a model of the evolution of inflectional classes, as they are reconstructed across generations via a combination of direct transmission and analogical inference. Parameters examined pertain to diversity of the lexicon and the cognitive policies governing inferential reasoning. We discover that persistence of stable inflection classes interacts in complex ways with features which affect how inflection classes are inferred. Thus we contribute to a greater understanding of factors affecting cultural attractors' existence, and to insights into a widespread and complex trait of human language.

## 1. Introduction

Human languages present an explanatory challenge: to reconcile their seemingly endless structural diversity with an equally striking tendency to exhibit recurrent similarities. Cultural evolutionary perspectives offer a solution. The adaptive flexibility of **cultural** transmission introduces the noise which makes linguistic innovation and change inevitable, while the **cognitive** nature of language influences the origin, direction and uptake of these changes, and their subsequent stability. Cultural Attraction Theory (Claidière & Sperber, 2007) foregrounds the role of cognition in shaping the adaptive landscape over which culture evolves. Cultural attractors are regions of meta-stability in cultural trait space. Their existence enables traits to remain stable in certain states but not others, thus promoting the emergence of shared, cumulative culture (see Griffiths & Kalish, 2007; Kirby, Tamariz, Cornish, & Smith, 2015). Here we examine *inflectional class systems* as cultural attractors. We investigate how parameters governing the inferential reconstruction of inflectional classes affect their evolutionary stability. Inflectional class systems are known to persist across millennia, thus the cognitive strategies common among language speakers are likely to be those which promote stability. Consequently, we view our approach as a method for identifying parameter settings of cognitive policies that may promote the persistence of a highly complex cultural attractor known to be a salient, recurrent trait of human language.

Table 1. Paradigms illustrating three inflectional classes in Swedish nouns.

| | | SG.INDEF | SG.DEF | PL.INDEF | PL.DEF |
|---|---|---|---|---|---|
| IC1 | 'school' | *skol**a*** | *skol**an*** | *skol**or*** | *skol**orna*** |
| | 'bottle' | *flask**a*** | *flask**an*** | *flask**or*** | *flask**orna*** |
| IC2 | 'chair' | *stol* | *stol**en*** | *stol**ar*** | *stol**arna*** |
| | 'box' | *ask* | *ask**en*** | *ask**ar*** | *ask**arna*** |
| IC3 | 'idol' | *idol* | *idol**en*** | *idol**er*** | *idol**erna*** |
| | 'Basque' | *bask* | *bask**en*** | *bask**er*** | *bask**erna*** |

**Inflectional Classes as Cultural Attractors** In many languages, lexemes can possess a set of multiple, contextually-conditioned wordforms, termed an *inflectional paradigm* as illustrated in Table 1 for Swedish nouns. In paradigms, each wordform occupies a *cell*. The wordforms in a paradigm typically contain formal differences, termed the *exponents* of the cell, and Inflectional Classes (ICs) are distinct *patterns* of exponents found in paradigms. In Table 1, IC 1 has exponents {*-a,-an,-or,-orna*} whereas IC 2 has {Ø,*-en,-ar,-arna*}.

As languages change, ICs are known to serve as attractors in the sense that anomalous paradigms (which may arise as the result of minor disruptions) often undergo alteration, to become more like other ICs in the language (Maiden, 2018). Recently, an explanation for this has been formulated in terms of the reconstruction and transmission of paradigms.

**Reconstruction and Transmission** Paradigms are reconstructed not *en bloc* but in piecemeal fashion, one cell at a time: a single utterance might contain the wordform '*sang*', but not the whole paradigm '*sing–sings–singing–sang–sung*'. In many languages, paradigms are sufficiently large (often comprised of dozens if not hundreds of wordforms) that it is implausible that speakers would always have heard a specific wordform of a lexeme before they first need to produce it (Ackerman, Blevins, & Malouf, 2009; Blevins, Milin, & Ramscar, 2017). Yet speakers can perform this task and they agree upon the solution. Psycholinguistic results show that speakers have clear inferential intuitions about paradigm forms they have not previously heard. English speakers for instance nominate *splung* but not *splong* as a plausible past tense form of novel *spling* (Albright & Hayes, 2002). Thus, cognitively, ICs are supported by processes of inferential reasoning in addition to mere storage of wordforms.

**Conditional Predictability** In natural languages, the distribution of exponents in different cells is not random, rather the exponent of one cell is typically relatively predictable from the exponent of another (Carstairs-McCarthy, 1987); conditional entropy is low compared to random covariation (Ackerman & Malouf, 2013). Linguists have been interested in understanding this interpredictability as a potentially emergent property of IC evolution (Ackerman & Malouf, 2015; Round, Beniamine, & Esher, 2021).

## 2. Modelling of Attractor Evolution

**Inflectional Classes** The task of inferring the content of a paradigm cell is known as the *paradigm cell filling problem*, or PCFP (Ackerman et al., 2009). Ackerman and Malouf (2015) modelled an analogical PCFP process, in which one cell of a lexeme's paradigm (which we call the *focal* cell) is inferred from analogical relationships that exist between it and one other cell (which we call the *pivot* cell, and which is sampled at random). To infer the focal cell's exponent, the agent attends to other lexemes that share a similar pivot cell, and samples from their focal cells. In an iterated learning simulation, with a single agent who learns all forms of the language except for one, which is inferred analogically, a system that is seeded initially with random paradigms will gradually self-organize into a limited number of ICs. Ackerman and Malouf (2015) show that across the lexicon, the mean conditional entropy of pairs of cells falls over time, and interpret this as reflecting a rising interpredictability between cells.

However, Round et al. (2021) note that conditional entropy can fall for two reasons: because interpredictability is increasing or because total entropy is falling as a system becomes uniform. They show that in Ackerman and Malouf (2015)'s model, conditional entropy falls for the latter reason: unlike the stable distinctiveness found in real IC systems, ICs in the model are inherently unstable and inevitably collapse together entirely. They advocate an improved measure of interpredictability based on mutual information. Mutual information will rise and fall as interpredictability does, including when total entropy is changing.

**Inferential Policies** In Ackerman and Malouf (2015) and Round et al. (2021), the PCFP is based solely on the contents of lexemes' inflectional paradigms. However, other information sources, including similarities along semantic and phonological dimensions, are known to affect the probability with which one lexeme influences another, both psycholinguistically and in language change (Hayes, Zuraw, Siptár, & Londe, 2009; Maiden, 2020). Here we investigate the emergence and stability of inflectional classes when analogical inference in the PCFP makes reference to *extra-paradigmatic similarity*: features other than morphological exponents that lexemes can share.

## 3. Model Description

We examine conditions for the (in)stability of cultural attractors, namely inflectional classes, in a multi-agent model in which paradigms evolve via a PCFP mechanism.[1] An initial population of agents is created who all share the same

---

[1] Model code and a Wiki guide are available at bit.ly/ELXIV.

lexicon. In each iteration, a child population is created, all of whom begin life with an empty lexicon. During acquisition, children learn from adults, who utter wordforms including by using the PCFP (see below) to infer forms which the adults themselves have not heard. Finally, adults die, children become adults, and the next iteration begins with the creation of a new child population.

Simulations commence with an initial population of $A$ adult agents, who share a randomised lexicon of $L$ lexemes[2] $\{\text{LEX}_1, \text{LEX}_2, \ldots, \text{LEX}_L\}$. Each lexeme has a paradigm of $C$ cells, whose exponents are represented as integers drawn from the range $\{1,2,\ldots V\}$. The initial lexicon contains $I_{init}$ distinct inflection classes which may share some (but not all) exponents. The model runs for $G$ iterations ('generations'). At each iteration, children during acquisition receive an input totaling $W$ wordforms, transmitted in equal measure by $Q$ randomly selected adult acquisition sources. If a child hears multiple wordforms for the same cell of a lexeme, it stores the first that it encounters. Adults sample wordforms to transmit as follows. A lexeme is sampled from among the lexemes known to the agent, according to a Zipfian distribution derived from the frequencies with which lexemes were heard by the agent as a child. A cell for that lexeme is chosen, again according to a Zipfian distribution. The exponent for the cell is retrieved from memory if the adult learned it as a child. Otherwise it is inferred by the PCFP process.

**The PCFP** The PCFP proceeds as follows. To infer the focal cell of the focal lexeme, (1) sample $E$ *evidence lexemes*; (2) accord them a weight of $+1$ for each non-focal cell that matches the corresponding cell of the focal lexeme; (3) for each exponent value $v$ in $\{1,2,\ldots,V\}$, assign it a weight equal to the summed weights of all evidence lexemes whose focal cell contains $v$. The selected exponent value is the one with the greatest weight; ties are broken randomly.

At step (1), the sampling of evidence lexemes is biased by lexemes' similarity in their extra-paradigmatic features. Moreover, the relative importance attached to each feature when calculating similarity is controlled by a feature *weight* which the agent *learns* at the conclusion of acquisition, as a function of the correlations it finds between a feature's values and lexemes' exponents.[3] As a language evolves, agents' feature weights—and the associated bias introduced into their PCFP solutions—will change over time, reflecting the shifting relationships between lexemes' unchanging extra-paradigmatic features and changing paradigms.

In this study, we investigate the effects of different kinds of correlations *in the initial population's lexicon* between extra-paradigmatic features and lexemes' ICs. Lexemes are given three features, allowing us to examine four conditions: *High*, in which one feature correlates exactly with ICs; *Medium*, in which one feature correlates perfectly with ICs for the first 50% of the lexicon; *Double medium*, which is identical to *Medium* but with the additional of a second feature that cor-

---

[2] In the remainder of this section, italicised single letters denote model parameters.

[3] Implemented using Random Forest Feature Importance. See model code for details.

relates perfectly with the ICs of the *second* 50% of the lexicon; and *Zero*, in which no feature correlates in these ways with initial ICs. Thus, in the *High, Medium* and *Double* conditions, whenever lexeme LEX_*i* is the subject of the PCFP, the set of evidence lexemes against which it is compared is more likely to be drawn from lexemes which shared LEX_*i*'s IC in the initial population's lexicon.

Table 2. Parameter definitions and values used. See main text for explanation.

| Parameter | | Value(s) | Parameter | | Value(s) |
|---|---|---|---|---|---|
| PARADIGMS | | | TRANSMISSION | | |
| Lexemes | $L$ | 200 | Agents | $A$ | 4 |
| Cells | $C$ | 5 | Acquisition sources | $Q$ | 3 |
| Possible cell values | $V$ | 5, 50 | Generations | $G$ | 5,000 |
| Inflection classes (initial) | $I_{init}$ | 5 | Wordforms in acquisition | $W$ | 24,000 |
| PCFP | | | | | |
| Evidence lexemes | $E$ | 10 | | | |
| Extra-paradigmatic bias | $B$ | High, Medium, Double, Zero | | | |

## 4. Simulation Experiments and Results

Parameter settings investigated are shown in Table 2. Varied parameters were $B$ (extra-paradigmatic bias) and $V$ (range of available exponents), taking 4 and 2 values respectively for a total of 8 conditions. Each condition ran 30 times.

The outcomes of primary interest are the change in number of ICs, the degree to which different agents conform in the ICs in their grammars, and the inter-predictability between cells in agents' grammar. To examine these we measure five population-wide statistics every 50 generations: (1) *IC Diversity*, the total number of distinct ICs in the grammars of the population of agents (lexemes with missing wordforms are ignored); (2) *IC Disparity*, defined as $1-A/D$ where $A$ is the average number of distinct ICs for individual agents and $D$ is IC Diversity; (3) *Conditional entropy*, the mean conditional entropy between pairs of cells in an agent's paradigm system, averaged across all agents; (4) *Mutual information*, the mean mutual information between pairs of cells in an agent's paradigm system, averaged across all agents; (5) *Maximum prediction weight*, the weight of the feature with respect to which the agent's selection of evidence lexemes is most strongly biased. Figure 1 visualises the four measures across 5,000 generations.

The results show initial, transient spikes in IC Diversity, IC Disparity and conditional entropy as the system departs from its initialisation state (first-generation agents share the same lexicon). Then, under all conditions, we find broad falls in IC Diversity and IC Disparity, falls in conditional entropy, and falls in mutual information as the newly-formed ICs begin to collapse together. The analogising effect of the PCFP is bringing different lexemes into line with one another. In almost all cases, each statistic takes a larger value in the 50-exponent case (blue lines). More exponents entail greater entropy and information because there

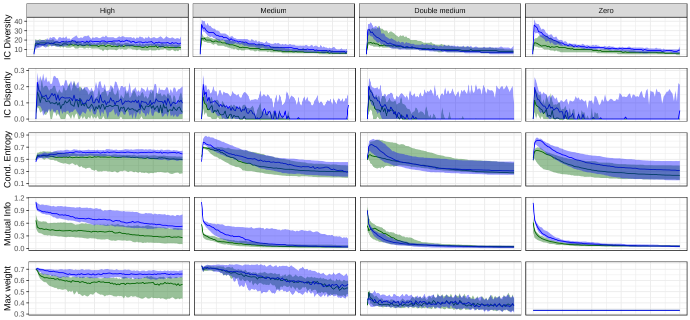


Figure 1. Five measures of ICs over 5,000 generations, observed at intervals of 50 generations. Conditions with 50 available exponents are in blue, with 5 in green. Each condition was run 30 times. Lines show medians, ribbons show 80% of variation. Each column of panels shows one bias level.

is a smaller chance of the same exponents being used across different ICs. And more exponents entail greater IC Diversity and Disparity because there is a greater source of variation with which to recombine existing ICs into new ones.

Differences in different bias conditions (i.e. across different columns in figure 1) can be understood as follows. First note all measures' uniqueness in the *High* condition. In *High*, evidence lexemes are very likely to be drawn from the same initial IC as the focal lexeme. Consequently, initial groups of lexemes will stick together through generations: they constitute reliable clusters exerting an internal gravitational pull, preventing individual lexemes from escaping to form new classes. Mutual information remains high because the initial ICs are mostly conserved. The other conditions show accelerated declines in conditional entropy and mutual information. Inflection classes are collapsing together across the lexicon; because the feature correlations are not strong enough to generate insular clusters, lexemes from different initial classes might be used to inform the PCFP for any given lexeme. Second, note the difference between the *Medium* and *Double* conditions. On the face of it, these patterns are paradoxical: the *Double* condition contains greater correlation across the lexicon, so one would expect its statistics to resemble those of the *High* condition. But its IC Disparity and mutual information are more similar to the *Zero* condition. How can more correlation across multiple features accelerate collapse, when maximum correlation in a single feature retards collapse? The answer is revealed by the 'Max weight' statistic, measuring how strongly an agent favours one set of extra-paradigmatic features over the others. In the *High* condition, agents favour the strongly-correlated feature. In the *Medium* condition, there is no strongly correlated column, but agents favour the medium-correlation feature. This leads to classes beginning to 'cluster' according to those features (even though they do not perfectly correspond to initial ICs), thus the collapse (decrease in IC Disparity and mutual information) is somewhat slowed.

By contrast, in the *Double* condition, **agents cannot strongly favour any single feature**; their assignment of lexemes to groups is pulled in two different directions indicated by the two distinct medium-correlation features. Two medium-strength correlations that pull in different directions cancel each other out; as a result, the statistics in this condition more resemble those of the *Zero* condition, where there is no information by which to favour certain lexeme clusters at all. These results indicate that IC stability may be affected in strikingly different ways by correlations which agents detect between inflection classes and extra-paradigmatic features, such as lexical semantics or phonological stem shape.

## 5. Conclusions and Future Directions

In this paper we presented a model of evolution in systems of inflectional classes, which are known to act as complex cultural attractors in languages around the world. Framing the study in terms of Cultural Attractor Theory, we modeled the reconstruction of ICs in terms of the paradigm cell filling problem (PCFP), which has attracted considerable recent attention in linguistics. We ran simulations chosen to highlight a key potential parameter of variation in the cognitive, inferential processes that underlie the PCFP and which we therefore expect to play a role in shaping the cultural attractor landscapes which ICs traverse. We find that multiple sources of analogical inference can prevent distinct categorisation, because those sources can conflict with each other. Therefore, explaining how agents categorise by analogy is not just a case of quantifying correlational relationships, but understanding the different sources of those correlations and their interactions.

Future elaborations of our approach include adding the effects of noise and channel biases during transmission, and transforming interaction into a **communication game** involving costs for communicative failure. More complex storage of heard exponents could allow agents to track the grammatical **variation** they encounter. Further possibilities are implemented in our model, but have not been explored in this study. Our model design enables agents to **converse as adults**, leading to paradigm gaps being filled via horizontal inheritance among a peer generation. Recent evidence (e.g. Raviv, Meyer, & Lev-Ari, 2020) underlines the importance of **population demographics**, e.g. population size and network structure (which our model implements), and overlapping, non-synchronous generations. Additional parameter values of the morphological system to be examined include **larger paradigms** with more lexemes, more cells, and multiple exponents per cell. Finally, it will be important to examine exogenous impacts on paradigm systems such as sound change and borrowing of lexicon from other languages; our model implementation anticipates the addition of these extensions.

## Acknowledgements

The authors wish to thank three anonymous reviewers for helpful feedback. This work was supported by British Academy Global Professorship GP300169 to ER and British Academy Newton International Fellowship NIF23\100218 to SB.

## References

Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 54–82). Oxford: Oxford University Press.

Ackerman, F., & Malouf, R. (2013). Morphological Organization: The Low Conditional Entropy Conjecture. *Language*, *89*(3), 429–464.

Ackerman, F., & Malouf, R. (2015). The No Blur Principle Effects as an Emergent Property of Language Systems. In *Proceedings of the 41st Annual Meeting of the Berkeley Linguistics Society* (pp. 1–14). Berkeley: BLS.

Albright, A., & Hayes, B. (2002). Modeling English past tense intuitions with minimal generalization. In *Proceedings of the ACL-02 workshop on morphological and phonological learning* (pp. 58–69). Philadelphia: Association for Computational Linguistics.

Blevins, J. P., Milin, P., & Ramscar, M. (2017). The Zipfian Paradigm Cell Filling Problem. In F. Kiefer, J. Blevins, & H. Bartos (Eds.), *Perspectives on Morphological Organization* (pp. 139–158). Leiden: Brill.

Carstairs-McCarthy, A. (1987). *Allomorphy in Inflexion.* London: Croom Helm.

Claidière, N., & Sperber, D. (2007). The role of attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1-2), 89–111.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive science*, *31*(3), 441–480.

Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009). Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language*, *85*(4), 822–863.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Maiden, M. (2018). *The Romance Verb: Morphomic Structure and Diachrony.* Oxford: Oxford University Press.

Maiden, M. (2020). The verbs ‘rain’ and ‘snow’ in Gallo-Romance, and other morphological mismatches in diachrony. In *Variation and Change in Gallo-Romance Grammar.* Oxford: Oxford University Press.

Raviv, L., Meyer, A., & Lev-Ari, S. (2020). The role of social network structure in the emergence of linguistic structure. *Cognitive Science*, *44*(8), e12876.

Round, E. R., Beniamine, S., & Esher, L. (2021). The role of attraction–repulsion dynamics in simulating the emergence of inflectional class systems. In *International Symposium of Morphology 2021.* (http://nabil.hathout.free.fr/ISMo2021)

# GROUP BINDING AS THE MAIN DRIVER OF LANGUAGE EVOLUTION

Daniil Ryabko$^{*1,2}$ and Alvaro Mbeju Moreno$^{2}$

$^{*}$Corresponding Author: daniil@ryabko.net
$^{1}$INRIA, France (on leave from)
$^{2}$Fishlife Research, Belize

A novel hypothesis concerning language evolution is advanced. It posits that languages have evolved as a means of binding individuals to a group, as well as for defining these groups. The key evolutionary adaptation is the loss, in adults, of the ability to learn languages with the ease that children possess. The proposed hypothesis helps to explain the diversity of human languages, their complexity, as well as the apparent uniqueness of the language faculty to humans.

## 1. The main hypothesis

The problem of language evolution has fascinated researchers for centuries (Wang and Minett, 2005), with the question of language diversity being one of the most intriguing ones (Lupyan and Dale, 2016). Why do peolple speak different languages? Whereas for groups of people that comminucate little or not at all some languages differences are expected to appear due to the effects akin to genetic drift, the langauge divergence between groups that live next to each other is much harder to explain (Labov, 2010). Moreover, unlike in biological reproduction where each new individual inherits genetic material from one or two parents, each new carrier of a language samples from a much larger set (dozens or even hundreds of carriers). Thus, it may be expected that mutations would average out and the new learner would adopt the norm that is most common; according to Nettle (1999), this problem in modelling language evolution has been recognized since at least Sapir (1921).

The importance of language in the individual's social identity is well-established, and is manifest in children from an early age (Kinzler, 2021; Cohen, 2012). We suggest that the need for social identity has been the main driver of the evolution of language. More specifically, language evolved in such a way as to make it difficult to learn as an adult. Consider the fact that children have the ability to learn a language perfectly and the adults do not, in the sense that it is extremely difficult or often impossible for an adult to learn to speak a language as well as someone who learned it as a child. The cornerstone question that we pose is the following:

*Can the  loss of the language-learning ability in adults be adaptive?*

In other words, can the loss of this ability serve some purpose, and if so then is it possible that languages have evolved to fulfill this purpose? Put this way, the question may appear baffling, since it is difficult to imagine what advantage a loss of a cognitive ability may confer to an individual. However, the advantages become clear if this question is considered on the level of groups. The loss of language-learning ability defines, for each language, the group of its native speakers, and effectively binds them to this group.

## 2. Altruism and group-level selection

Binding individuals to a group allows for the evolution of altruistic behaviour via group-level selection, which means that the evolution of language has to be considered in the context of the evolution of altruism. Let us consider this latter point in more detail.

Group-level (or, more generally, multilevel) selection is used to explain the evolution of altruistic behaviour. Altruistic behaviour is a behaviour that is detrimental to the individual but is beneficial to others which may be only distantly related; non-altruistic behaviour is called selfish. If selection acts on the super-individual level, that is, if groups compete between themselves, then groups which have more altruists get an advantage, even though an altruist is disadvantaged within their group (so long as it has at least one selfish member). It is worth noting that the idea of group-level selection as an explanation for the emergence of altruistic behaviour can be traced back to Darwin (1871).

In sociolinguistics, it is generally acknowledged that language should be considered a property of a group first and individual only second. In the words of Labov (2010), who calls is the central dogma of sociolinguistics, language is seen as an abstract pattern located in the speech community and exterior to the individual. The function of language as a source of social markers is widely recognized and well-studied (Nettle and Dunbar, 1997; Roberts, 2013). Moreover, the diversity of language seems interfere with the its other functions, such as communication (Labov, 2010; Roberts, 2013). Cohen (2012) gives a list of properties of language (more specifically, of accent) that makes it a particularly reliable source of social markers. Of particular importance for the hypothesis proposed is that it is salient and hard to fake. Indeed, one's native accent is hard to impossible to hide in a foreign language (Cohen, 2012; Sigmund and Nowak, 2001; Roberts, 2008), and it is extremely difficult for an adult to learn to speak a foreign language as well as a native speaker (Birdsong, 2009; Roberts, 2013; Kinzler, 2021).

Mathematical and computational models of various complexity have been proposed to explain the evolution of altruism through group-level selection, and, more specifically, tag-based cooperation; see (Rand and Nowak, 2013) for a review. Again, for the purpose of the proposed the hypothesis the most important aspect is the need for a mechanism that binds individuals to groups. Intuitively, it is clear that between-group migration works against cooperation; specifically, one might think of individuals fleeing a group in the face of danger, e.g. under attack by another group. Mathematical modelling not only confirms this intuition, but shows that already random migration has this effect. One of the simplest models that allows for migration is that by Traulsen and Nowak (2006). In this model, the population is divided into groups and individuals reproduce within the groups. Groups may split upon reaching a certain size, and individuals may migrate at random with a certain probability. The authors show that group-level selection favours altruism if $b/c > 1 + z + n/m$, where $b/c$ is the ratio between the benefit and cost of the altruistic act, $n$ and $m$ denote the maximum group size and the number of groups, and, finally, $z$ is the average number of migrants arising from one group during its lifetime. Thus, at least in this model, migration works directly against altruism: the higher the migration ($z$) the bigger the benefit of altruism should be related to its cost ($b/c$) in order for altruism to survive.

Group-level selection in our own species had almost certainly taken place; in particular, Tattersall (2016) notes that the spread of *Homo sapiens* over the planet between 100kyr and 50kyr ago occurred far too rapidly to be accounted for by the slow workings of natural selection at the individual level (and suggests an explanation based on the invention of language, but does not consider its role as a group-binding tool).

## 3. Other mechanisms of group-binding

If binding individuals to groups is an essential part of group evolution, and in particular of the evolution of altruism, then one should expect there to be more than one means of achieving it. Some examples of the manifestations of this phenomenon can be found in artificial body deformations. These have been found across a wide variety of human cultures, and include cranial deformations, foot-binding, scarification, tattooing and so on, which can often be linked to class, status, religion or other forms of social identity (Mackie, 1996; Hoshower et al., 1995; Ludvico and Kurland, 1995). While it is clear that these phenomena are cultural, their widespread presence in only distantly related populations is an evidence of an underlying common genetic adaptation. It is worth noting that many of the artificial body deformations mentioned are applied to young children, i.e., at the same age that language acquisition starts; others are applied at adolescence, where group membership is being ascertained (e.g., as a right of passage).

The presence of these group-binding practices provides an indirect support to the hypothesis advanced in this note.

One can also ask the question of whether there is anything that makes the language a better group-binding tool than these practices. One possible answer it the universality of language: everyone has it. Indeed, Cohen (2012) lists this property as one of those making the accent an especially good source of social markers. While some of the other social practices mentioned above may be more flexible, none is based on something that readily exists in all humans.

## 4. Discussion

We have proposed a hypothesis that language has a evolved as a tool to bind individuals to groups; thus, the difficulty with which adults can learn it is a group-level evolutionary adaptation. The need for languages to be complex is a direct consequence of this hypothesis (they need to be complex so as to be difficult for adults to learn). Moreover, for group-level selection to take place, groups needs to reproduce, which entails language diversity (languages diverge to accommodate or to define group splits). Thus, the proposed hypothesis fits well the empirically observed phenomena – the complexity and diversity of languages – that have so far remained largely enigmatic from the theory point of view. Moreover, it is supported by theoretical models of group evolution as well as by the evidence of other group-binding practices that permeate human populations around the world.

There are many questions left to be answered. Most importantly, while the proposed hypothesis attempts to explain why the language evolved and what purposes it serves, it remains so far to speculate exactly how this evolutionary process took place. One can envisage two distinct possibilities: either the language evolved on the basis on some pre-existing cognitive abilities that children possess and that they lose growing up; or, the evolutionary process has somehow suppressed some existing cognitive ability; or, which is most likely, it was some combination of the two. If some pre-existing abilities played the main role, we would perhaps be able to find some other cognitive abilities that children have but adults do not. These appear rather difficult to find in humans. However, looking beyond our own species, a related mechanism can be found, namely that of imprinting. It is widespread in the so-called precocial species, i.e. those in which the young are born relatively mature. Specifically, filial imprinting establishes social preference for an object that should typically be the parent, during the first hours or days of life; sexual imprinting works similarly to establish sexual preference (see, e.g., Reznikova (2007) for review). Most importantly, these mechanisms are only active during certain sensitive periods of life. Clearly, humans are not precocial species and do not need to imprint the parental image for the same reasons that gees chicks do. However, similar cognitive mechanisms could be present. Sexual imprinting is also interesting to consider from this point of view, as the choice of sexual partner is largely affected by their group membership and social standing.

Parallels can be drawn with language acquisition and change. Thus, children are sensitive to language variation from an early age, and they start to use it is a strong source of social markers (Johnson and White, 2020; Kinzler, 2021); for example, 5-6 year-old children were demonstrated to prefer native-accented speackers and this preferences was stronger than that for race (Kinzler et al., 2009). However, it is during the adolescence that the propensity to make changes to the language is the strongest Eckert (2004); Kinzler (2021). It thus may be interesting to study the emergence of new language forms from the point of view of reproduction of groups and its relation to the sexual sexual partner choice.

As with any hypothesis, an important question to consider is what evidence would disprove it. Finding other compelling reasons for the loss of the language-learning ability in adults, corroborated by studies on related species would be one way of doing it. Finding other learning abilities that have nothing to do with language (so that it could not be argued that language has developed on their basis) that disappear in adulthood would considerably undermine the hypothesis. It could also be useful to consider under which circumstances a language can actually become simpler. One example, that may appear contradicting the hypothesis proposed, is that languages spoken by larger with many non-native speakers have a tendency to lose some morphological complexity (Lupyan and Dale, 2010). However, it can be argued that precisely the fact that there are many non-native speakers (and that the is spoken by a large group) weakens its role as a source of social identity. In such cases this role is perhaps relegated to dialects, accents and other local language variations.

Another question that acquires a new perspective in light of the hypothesis advanced is the exclusivity of the capacity for language to our species. Language is often considered a uniquely human ability (e.g. Jackendoff, 2011), or at least partially so, in particular, Hauser et al. (2002) suggest that the FLN is uniquely human. From the perspective of the evolution of groups and altruism, our species does appear to be unique, though by no means the most advanced. Humans are perhaps the only species to live in large, highly-structured but at the same time genetically heterogeneous communities. Eusocial species, such as bees, ants and mole rats, are more advanced in the sense that their communities are much more genetically homogeneous. Some of these latter species, in particular bees and a select few species of ants, possess communication systems that allow for transmission of quantitative information, and, at least in the case of ants, are flexible enough to allow for information compression (review in Reznikova and Ryabko, 2011). Since they are eusocial, the groups in these species are families, i.e., are bound genetically, and so do not need to bind their members further using language or other means (this said, in many ant species the colonies are actually groups of families, and thus are genetically heterogeneous). Therefore, if the hypothesis advanced in this note is correct, their languages should be much simpler and less diverse. While their relative (with respect to humans') simplicity is perhaps a foregone conclusion, the diversity aspect appears an interesting subject for a comparative study.

## References

Birdsong, D. (2009). Age and the end state of second language acquisition. *The new handbook of second language acquisition*, 17(1):401–424.

Cohen, E. (2012). The evolution of tag-based cooperation in humans: The case for accent. *Current Anthropology*, 53(5):588–616.

Darwin, C. (1871). *The descent of man: and selection in relation to sex*. John Murray, Albemarle Street.

Eckert, P. (2004). Adolescent language. *Language in the USA: themes for the twenty-first century*, 6:58–75.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579.

Hoshower, L. M., Buikstra, J. E., Goldstein, P. S., and Webster, A. D. (1995). Artificial cranial deformation at the omo m10 site: A Tiwanaku complex from the Moquegua valley, Peru. *Latin American Antiquity*, 6(2):145–164.

Jackendoff, R. (2011). What is the human language faculty? two views. *Language*, pages 586–624.

Johnson, E. K. and White, K. S. (2020). Developmental sociolinguistics: Children's acquisition of language variation. *Wiley Interdisciplinary Reviews: Cognitive Science*, 11(1):e1515.

Kinzler, K. D. (2021). Language as a social cue. *Annual review of psychology*, 72:241–264.

Kinzler, K. D., Shutts, K., DeJesus, J., and Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social cognition*, 27(4):623–634.

Labov, W. (2010). Principles of linguistic change, cognitive and cultural factors. *Malden, MA: John Wiley & Sons*, 10:9781444327496.

Ludvico, L. R. and Kurland, J. A. (1995). Symbolic or not-so-symbolic wounds: The behavioral ecology of human scarification. *Ethology and Sociobiology*, 16(2):155–172.

Lupyan, G. and Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, 5(1):e8559.

Lupyan, G. and Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends in cognitive sciences*, 20(9):649–660.

Mackie, G. (1996). Ending footbinding and infibulation: A convention account. *American sociological review*, pages 999–1017.

Nettle, D. (1999). Using social impact theory to simulate language change. *Lingua*, 108(2-3):95–117.

Nettle, D. and Dunbar, R. I. (1997). Social markers and the evolution of reciprocal exchange. *Current Anthropology*, 38(1):93–99.

Rand, D. G. and Nowak, M. A. (2013). Human cooperation. *Trends in cognitive sciences*, 17(8):413–425.

Reznikova, Z. (2007). *Animal intelligence: From individual to social cognition.* Cambridge University Press.

Reznikova, Z. and Ryabko, B. (2011). Numerical competence in animals, with an insight from ants. *Behaviour*, 148(4):405–434.

Roberts, G. (2008). Language and the free-rider problem: An experimental paradigm. *Biological Theory*, 3(2):174–183.

Roberts, G. (2013). Perspectives on language as a source of social markers. *Language and Linguistics Compass*, 7(12):619–632.

Sapir, E. (1921). *Language, race and culture.* Harcourt Brace & Company.

Sigmund, K. and Nowak, M. A. (2001). Tides of tolerance. *Nature*, 414(6862):403–405.

Tattersall, I. (2016). A tentative framework for the acquisition of language and modern human cognition. *Journal of Ahropological Sciences*, 94:157–166.

Traulsen, A. and Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences*, 103(29):10952–10955.

Wang, W. S. and Minett, J. W. (2005). The invasion of language: emergence, change and death. *Trends in ecology & evolution*, 20(5):263–269.

# HUMAN LANGUAGE LAWS IN CHIMPANZEE SEXUAL SOLICITATION GESTURES

ALEXANDRA SAFRYGHIN[1], CATHARINE CROSS[1], BRITTANY FALLON[1], RAPHAELA HEESEN[3], RAMON FERRER-I-CANCHO[2&], and CATHERINE HOBAITER[1,4&*]

*Corresponding Author: clh42@st-andrews.ac.uk (CH)
[1]School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK
[2]Departament de Ciències de la Computació, Universitat Politècnica de Catalunya, Barcelona, Spain
[3]Department of Psychology, Durham University, Durham, UK
[4]Budongo Conservation Field Station, Masindi, Uganda
[&] These authors contributed equally to this work

## 1. Introduction

Two language laws have been identified as manifestations of universal principles of animal behaviour (Semple et al., 2021). Zipf's law of brevity describes a negative relationship between the length of a behaviour and the frequency of its use (Zipf, 1949). Menzerath's law describes a negative correlation between the number of behaviours in a sequence and the average length of behaviours composing it (Menzerath, 1954). Following Torre et al., 2019's physical hypothesis on the emergence of language laws and given the key role gestural communication played in human language evolution (*review:* Rodrigues et al., 2021), chimpanzee gestural communication represents a powerful model in which to explore compression and language laws (Ferrer-i-Cancho et al., 2022), with repertoires of over 70 distinct gesture types (Byrne et al., 2017). Menzerath's law appears to hold in play gesture sequences, but the play repertoire represents a rare failure of Zipf's law, perhaps due to the nature of that specific context (Heesen et al., 2019). Here, we test Zipf's law of brevity and Menzerath's law in male chimpanzee sexual solicitation gestures, which, in contrast to play, are subject to strong selection pressures for success (Hobaiter & Byrne, 2012).

## 2. Methods

We measured 560 male-to-female sexual solicitation gestures from 173 videos of 16 wild, habituated East African chimpanzees (*Pan troglodytes schweinfuthii*) from the Sonso community, Budongo Forest Reserve, Uganda. We calculated gesture duration following Heesen et al. (2019). Gestures were grouped into *types,* with each instance being an individual *token* (see supp mat). Tokens performed with less than 1s between them formed a sequence of *n* tokens. Single gestures formed sequences of length 1. Following Heesen et al. (2019), we computed correlations and compression values related to the respective laws. We also performed subset analysis, grouping gesture types into manual and whole-body gestures.

## 3. Results

Zipf's law of brevity did not hold. A tendency towards an opposite pattern was detected (Spearman correlation - all data: $r_s=0.30$, $n=26$, $p=0.066$; manual subset: $r_s=0.42$, $n=21$, $p=0.031$; supp mat Figure 1) and was corroborated by the permutation analysis. The expected mean code length $L$ of the data tended to be significantly big, rather than small, compared to the distribution of $L$ calculated via $10^5$ permutations (all data: $L=0.239$s, $p_{right}=0.05$; manual subset: $L=2.26$s, $p_{right}=0.058$). For whole-body gestures no pattern was detected.

While Spearman's test revealed only a trend towards correlation between average gesture duration within sequence $t$ and sequence size $n$ (Spearman correlation: $r_s=-0.08$ $n=359$, $p=0.076$; supp mat Figure 2), the total sum of the duration of each sequence $M$ was significantly small ($M=1300.67$, $n=359$, $p=0.003$) suggesting a linear association between $n$ and $t$ following Menzerath's law.

## 4. Discussion

Our results challenge the view that compression is a universal principle in animal communication (Ferrer-i-Cancho et al., 2013; Börstell et al., 2016) as it does not act on a communicative system uniformly. Although 77% of gesture types described here matched those in play sequences (Heesen et al., 2019), and data belonged to the same individuals, collected over the same period, the two different contexts produced conflicting results: in sexual solicitations we detected an opposite pattern to Zipf's law of brevity and no clear evidence for Menzerath's law. In chimpanzee solicitation the benefits of successful communication to individual fitness appear to outweigh the energetic costs associated with the production of prolonged signals in such a highly competitive environment, fitting with theories on sexual selection of traits (Zahavi, 1975). Although an apparent absence of a pattern may depend on the unit of analysis (Demartsev et al., 2019), there is scope for investigating different parts of gesture structure other than duration, which may shed light on the physical or environmental mechanisms that led to the emergence of these patterns in modern human languages.

## Acknowledgements

We thank the staff and field assistants of the Budongo Conservation Field Station, the Ugandan National Council for Science and Technology and the Uganda Wildlife Authority for making this research possible. We thank the Royal Zoological Society of Scotland for its funding of the field station. This research received funding European Union's 8th Framework Programme, Horizon 2020, under grant agreement no 802719.

## References

Börstell, C., Hörberg, T., & Östling, R. (2016). Distribution and duration of signs and parts of speech in Swedish Sign Language. *Sign Language and Linguistics*, *19*(2), 143-196. doi: 10.1075/sll.19.2.01bor

Byrne, R., Cartmill, E., Genty, E., Graham, K., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Animal Cognition*, *20*(4), 755-769. doi: 10.1007/s10071-017-1096-4

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory*. 2nd ed. Hoboken, N.J: Wiley-Interscience.

Demartsev, V., Gordon, N., Barocas, A., Bar-Ziv, E., Ilany, T., & Goll, Y. et al. (2019). The "Law of Brevity" in animal communication: Sex-specific signaling optimization is determined by call amplitude rather than duration. *Evolution Letters*, *3*(6), 623-634. doi: 10.1002/evl3.147

Ferrer-i-Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M., & Semple, S. (2013). Compression as a Universal Principle of Animal Behavior. *Cognitive Science*, *37*(8), 1565-1578. doi: 10.1111/cogs.12061

Ferrer-i-Cancho, R. & Bentz, C & Seguin, C. (2022). Optimal coding and the origins of Zipfian laws. *Journal of Quantitative Linguistics*, *29*(2), 165-194. doi: 10.1080/09296174.2020.1778387

Heesen, R., Hobaiter, C., Ferrer-i-Cancho, R., & Semple, S. (2019). Linguistic laws in chimpanzee gestural communication. *Proceedings of The Royal Society B: Biological Sciences*, *286*(1896), 20182900. doi: 10.1098/rspb.2018.2900

Hobaiter, C., & Byrne, R. W. (2012). Gesture use in consortship: Wild chimpanzees' use of gesture for an 'evolutionarily urgent' purpose. In S. Pika & K. Liebal (Eds.), *Gesture Studies* (6), 129–146. John Benjamins Publishing Company. https://doi.org/10.1075/gs.6.08hob

Menzerath, P. (1954). *Die architektonik des deutschen wortschatzes.* F. Dümmler (Ed).

Semple, S., Ferrer-i-Cancho, R., & Gustison, M. (2022). Linguistic laws in biology. *Trends In Ecology & Evolution*, *37*(1), 53-66. doi: 10.1016/j.tree.2021.08.012

Rodrigues, E., Santos, A., Veppo, F., Pereira, J., & Hobaiter, C. (2021). Connecting primate gesture to the evolutionary roots of language: A systematic review. *American Journal Of Primatology*, *83*(9). doi: 10.1002/ajp.23313

Torre IG, Luque B, Lacasa L, Kello CT, Hernández-Fernández A. (2019). On the physical origin of linguistic laws and lognormality in speech. *Royal Society Open Science*, 6, 191023. doi:10.1098/rsos.191023

Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal Of Theoretical Biology*, *53*(1), 205-214. doi: 10.1016/0022-5193(75)90111-3

Zipf, G. K. (1949). *Human behavior and the principle of least effort: an introduction to human ecology.* Addison-Wesley Press.

# THE FUNCTION AND EVOLUTION OF CHILD-DIRECTED SPEECH

JOHANNA SCHICK*[1,5], CAROLINE FRYNS[2,5], FRANZISKA WEGDELL[1,5], MARION LAPORTE[7,8], KLAUS ZUBERBÜHLER[2,4,5], CAREL P. VAN SCHAIK[3,5], SIMON W. TOWNSEND[1,5,6], and SABINE STOLL[1,5]

*Corresponding Author: johanna.schick@uzh.ch
[1]Department of Comparative Language Science, University of Zurich, Zurich, Switzerland
[2]Department of Comparative Cognition, University of Neuchatel, Neuchatel, Switzerland,
[3]Department of Anthropology, University of Zurich, Zurich, Switzerland
[4]School of Psychology and Neuroscience, University of St. Andrews, St. Andrews, United Kingdom
[5]Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Zurich, Switzerland
[6]Department of Psychology, University of Warwick, Warwick, United Kingdom
[7]Histoire naturelle de l'Homme préhistorique, UMR 7194, Paleo FED, Muséum National d'Histoire Naturelle, Paris, France
[8]Institut des Sciences du Calcul et des Données, Sorbonne Université, Paris, France,

## Abstract

Humans exhibit unusual vocal behaviour when interacting with infants and small children. In this 'child-directed speech' (CDS) the speech signal differs from speech addressed to adults, both acoustically and structurally, exhibiting increased conspicuousness and comprehensibility compared to regular adult-directed speech. CDS has been observed in numerous cultures and it is widely considered a human universal (Ferguson, 1978; Fernald et al., 1989; Soderstrom, 2007).

Over the past few decades, a plethora of studies has shown that these structural and acoustic features support the acquisition of language by infants at both comprehension (e.g. Singh, Nestor, Parikh, & Yull, 2009) and production (e.g. Porritt, Zinser, Bachorowski, & Kaplan, 2014; Rowe, 2012) levels, suggesting a functional teaching role.

Here, we discuss how CDS evolved by examining which elements of human CDS are present in our closest-living relatives and thus presumably also in the last common ancestor. We are interested both in frequency of CDS, i.e., how often the different species address their infants directly and whether they deploy features which increase conspicuousness/ostention (e.g. repetitions, F0, pitch modulation).

Research on non-human great-apes suggests that apes rarely direct vocal behaviour at their infants (chimpanzees, *Pan troglodytes*: (Laporte & Zuberbühler, 2011); bonobos, *Pan paniscus*: (Oller et al., 2019)), and if so, most likely in ways indistinguishable from adult-directed communication (ongoing work). This

would then suggest that infant great apes mostly rely on infant-surrounding communication in the acquisition of their vocal communicative competence, and that infant-directed communication evolved *de novo* among our hominin ancestors. As such, we argue that child-surrounding speech in humans requires much more research attention to understand whether this originally predominant source of input remains significant in humans, or has been replaced by CDS (see Figure 1) as the main source of learning. The potential relevance of child-surrounding speech is supported by the fact that the interplay of child-directed and child-surrounding speech shows major cross-cultural variation in human child-rearing. Indeed, in some linguistic communities, surrounding speech seems to be the primary source of input in the early phases of child development since adults rarely directly address children (Kaluli and Samoan: (Ochs & Schieffelin, 1984); Yucatec Mayan: (Shneidman & Goldin-Meadow, 2012), Tsimane: (Cristia, Gurven, & Stieglitz, 2019)), at least in their first year of life. Despite these differences in input type, children still become competent native speakers (Brown, 2011; De León, 2011; Ochs & Schieffelin, 1984).

This raises critical questions regarding current theories and the importance of CDS compared to child-surrounding speech in the evolution of language. We conclude that, even though CDS does indeed serve to facilitate language acquisition in humans, it also remains plausible that child-surrounding speech can compensate for limited CDS, at least during some developmental stages, thus deserving more consideration in ontogenetic studies of language evolution.

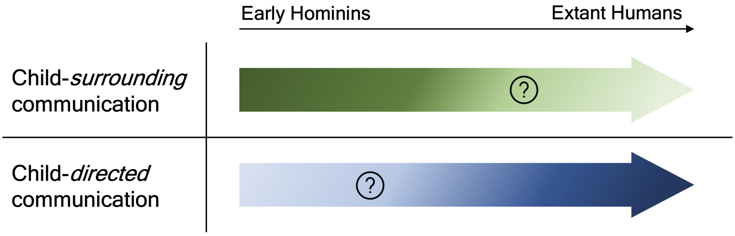


Figure 1. **Transition of child-surrounding to child-directed communication.** The transition of the importance of use of child-surrounding communication to child-directed communication. Darker colour shows importance/presence and brighter colour possible insignificance of CSS and CDS from early hominins to extant humans.

## References

Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook of language socialization* (pp. 29–55). Malden, MA: Wiley Online Library.

Cristia, A., Gurven, M., & Stieglitz, J. (2019). Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study. *Journal of Child Development*, *90*(3), 759–773.

De León, L. (2011). Language Socialization and Multiparty Participation Frameworks. In E. O. A. Duranti & B. B. Schieffelin (Eds.), *The handbook of language socialization.*

Ferguson, C. A. (1978). Talking to children: A search for universals. *Universals of human language*, *75*(4), 203–224.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. de, & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501.

Laporte, M. N., & Zuberbühler, K. (2011). The development of a greeting signal in wild chimpanzees. *Developmental Science*, *14*(5), 1220–1234.

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socializations: Three developmental stories and their implications. In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self and emotion* (p. 276–320). Cambridge: Cambridge University Press.

Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., & Call, J. (2019). Language Origins Viewed in Spontaneous and Interactive Vocal Rates of Human and Bonobo Infants. *Frontiers in Psychology*, *10*.

Porritt, L. L., Zinser, M. C., Bachorowski, J.-A., & Kaplan, P. S. (2014). Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Language Learning and Development*, *10*(1), 51–67.

Rowe, M. L. (2012). A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. *Child Development*, *83*(5), 1762–1774.

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673.

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, *14*(6), 654–666.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532.

# THE PUZZLE OF GREAT APE GESTURE & A SOLUTION

THOM SCOTT-PHILLIPS[1,*] and CHRISTOPHE HEINTZ[1]

[1] Department of Cognitive Science, Central European University
*Corresponding Author: thom.scottphillips@gmail.com

Searching for evolutionary homologies between humans and non-human great apes, one promising approach is to focus on social cognition and in particular in cognitive capacities for attention manipulation. Indeed, arguably the most intriguing thing about great ape interaction is how they often inform one another in ways that can seem very 'human' (for an example see e.g. Genty & Zuberbühler, 2014). *How to describe great ape gesture in a way that accounts for both its behavioural similarity with some forms of human communication, but also its limited range relative to the human case?* One proposed solution to this puzzle focuses on the role of 'we intentionality' in human communication (Tomasello, 2008). In our view we-intentionality is not a cognitive process but a behavioural phenomenon itself in need of explanation. Here we summarise a novel approach.

Figure 1 presents five embedded subsets, each a more or less specific means of manipulating others' attention (from Scott-Phillips & Heintz, in press). This 'special case of' approach is a novel alternative to existing frameworks for animal communication that are based on a supposed hierarchy of 'levels of intentionality' (see e.g. Townsend et al., 2017). It also helps to meet a growing demand for analyses that go beneath the surface of behaviour to consider underlying cognitive processes (e.g. Graham et al., 2020).

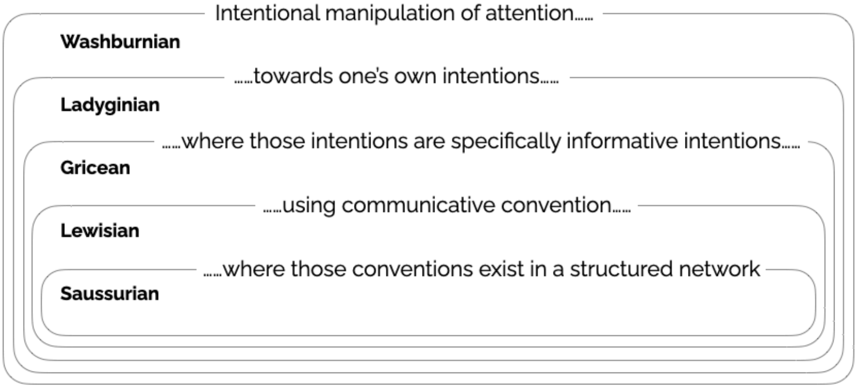


Figure 1. *Cognitive means of attention manipulation*. See main text for detailed description.

To describe these subsets we begin in the middle. With ‘Gricean’ we follow established characterisations in the Relevance Theory literature (e.g. Wharton, 2008; Wilson & Sperber, 2012; *inter alia*). This is the intentional manipulation of attention towards one’s own informative intentions. Working inwards, ’Lewisian’ is Gricean communication by means of convention (following Lewis, 1969). A great many instances of Gricean communication entail communicative conventions, such as nodding, winking and pointing. At the same time, humans can improvise non-conventional communication when necessary, sometimes called ‘pantomime’ (e.g. Zlatev et al., 2020). A further subset of Lewisian communication is when the conventions in question are organised in structured networks, commonly called ‘languages’. We call this layer ‘Saussurian’, following the emphasis that Ferdinand de Saussure placed on structure in linguistic analysis. At the other extreme, the outermost subset includes all instances of the intentional manipulation of attention. We assume this subset extends well beyond humans. We label it ‘Washburnian’ after Margaret Floy Washburn, who pioneered the study of cognition in a wide variety of non-human animals, at a time when more behaviourist approaches were coming into vogue (Washburn, 1908).

The key novelty in our analysis is specification of the second innermost subset (see also Warren & Call, 2022, who present a somewhat similar analysis). Here, individuals intentionally manipulate others’ attention towards evidence of the intended outcome, which could be, for instance, grooming, play or sex. This is possible if observers can identify such intentions on the basis of the observed behaviour. We name this subset ‘Ladyginian’ after Nadezhda Ladygina-Kohts (born Nadezhda Ladygina), who was one of the first scientists to document in a systematic way the similarities and differences between the expressive behaviours of humans and chimpanzees. The difference with Washburnian is that whereas Washburnian behaviour is intentional towards any ends, Ladyginian behaviour intentionally reveals intentions (to be groomed, to play, etc). The difference with Gricean is that Gricean behaviour intentionally reveals not simply intentions but specifically informative intentions.

We suggest that most great ape gesture may be best characterised as Ladyginian but not Gricean. This suggestion is potentially convergent with other, previous analyses of great ape gesture (in particular Moore, 2016; 2017). However, those previous analyses maintain the label ‘Gricean’. This conflates two distinct modes of interaction that can and, in our view, should be separated. Explaining the evolutionary origins of human communication, and hence languages, requires identification of the ecological reasons why humans, and apparently only humans, transitioned from Ladyginian to Gricean modes of interaction (Heintz & Scott-Phillips, in press).

## Funding statement

TSP & CH were financially supported by the European Research Council, under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 609819 (Somics project).

## References

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: Intentional communication with a rich set of innate signals. *Animal Cognition, 20*(4), 755-769.

Genty, E., & Zuberbühler, K. (2014). Spatial reference in a bonobo gesture. *Current Biology, 24*(14), 1601-1605.

Graham, K. E., Wilke, C., Lahiff, N. J., & Slocombe, K. E. (2020). Scratching beneath the surface: intentionality in great ape signal production. *Philosophical Transactions of the Royal Society B, 375*(1789), 20180403.

Heintz, C., & Scott-Phillips, T. (in press). Expression unleashed: The evolutionary & cognitive foundations of human communication. *Behavioral & Brain Sciences*.

Lewis, D. (1969). *Convention*. Harvard University Press.

Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Animal Cognition, 19*(1), 223-231.

Moore, R. (2017). Convergent minds: Ostension, inference and Grice's third clause. *Interface Focus, 7*, 20160107.

Scott-Phillips, T., & Heintz, C. (in press). Animal communication in linguistic & cognitive perspective. *Annual Review of Linguistics*.

Tomasello, M. (2008). *Origins Of Human Communication*. MIT Press.

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., Braga Goncalves, I., Burkart, J. M., Flower, T., Gaunet, F., Glock, H. J., Gruber, T., Jansen, D. A. W. A. M., Liebal, K., Linke, A., Miklósi, Á., Moore, R., van Schaik, C. P., Stoll, S., Vail, A., Waller, B. M., Wild, M., Zuberbühler, K., & Manser, M. B. (2017). Exorcising Grice's ghost: An empirical approach to studying intentional communication in animals. *Biological Reviews, 92*(3), 1427-1433.

Warren, E., & Call, J. (2022). Inferential communication: bridging the gap between intentional and ostensive communication in non-human primates. *Frontiers in Psychology, 12*, 718251.

Washburn, M. F. (1908). *The Animal Mind: A Text-Book of Comparative Psychology*. Macmillan Company.

Wharton, T. (2008). "Meaning$_{NN}$" and "showing": Gricean intentions and relevance-theoretic intentions. *Intercultural Pragmatics, 5*(2), 131-152.

Wilson, D., & Sperber, D. (2012). *Meaning & Relevance*. Cambridge University Press.

Zlatev, J., Żywiczyński, P., & Wacewicz, S. (2020). Pantomime as the original human-specific communicative system. *Journal of Language Evolution, 5*(2), 156-174.

# “MUSIC” PRODUCTION BY COCKATIELS

YOSHIMASA SEKI [*1]

[*]Corresponding Author: yoshimasa.seki@gmail.com
[1]Department of Psychology, Aichi University, Toyohashi, Japan

Elucidating the biological substrates essential for music (*i.e.,* musicality) could increase our understanding of the evolution of human language. Researchers have done work towards this aim using both humans and non-human animals; however, defining the concept of “music” is difficult and discussion of the evolution of musicality has not yet converged on a single idea (Savage et al., 2021). Some birdsongs may be similar to human music (Baptista & Keister, 2005); however, this does not mean these songs are universally considered to be music. Further, studies on music *production* (or, imitation) by non-human animals are very scarce, though some avian species (*e.g.,* bullfinches, starlings, Amazon parrots and African Greys) occasionally imitate human music.

Therefore, I documented four types of sound production behaviors (similar to human music production) in 3 cockatiels. First, (1) cockatiels exposed to a melody of a popular song (so it was universally considered music) spontaneously imitated the melody without any food reinforcement. Then, (2) the cockatiels spontaneously sang the song in synchrony with a playback of the melody; (2*a*) when the melody was played back shortly after a bird started singing the song, the bird paused singing and resumed after a short period to synchronize the timing of his singing to that of the playback melody; (2*b*) when the melody was played back while a bird was not vocalizing, the bird started singing and skipped some initial notes to synchronize the timing of his singing to that of the playback melody. So far, only a few studies have reported unison-like singing by several wild songbirds; however, the present examples might be a striking demonstration of this because the cockatiels sang a melody of human music in unison. Further, a few months later, (3) the birds spontaneously rearranged (or, customized) the melody. Some of them inserted novel sound elements around the tail of the melody. Finally, (4) they later produced novel rhythmic sound sequences which did not share similarities with sounds which occurred in their living environment, which means they created original sound sequences. The motor commands to produce these sounds were not likely inherent in this species because the acoustic patterns varied greatly among the individuals. Further, one of the birds produced a sound sequence by utilizing both vocal sounds and sounds generated by drumming a food cup with his beak (*see*, Le Covec et al., 2019). This is similar to a behavior observed in Palm cockatoos (Heinsohn et al., 2017). However, the cockatiel combined learned vocalizations (an imitation of a human word) and the sound made by hitting his beak against a hard surface, and repeated them in regular short intervals (similar to a human song that contains both vocals and drumming).

The results suggest that the cockatiels and humans share some of the capabilities necessary for music production. Because we only have a few reports on sound production behavior in wild cockatiels, it is difficult to discuss how the findings are involved in their ecology. As this report demonstrates, captive cockatiels are prominent imitators of man-made music; however, interestingly, they are not known to be proficient at imitating long stretches of human speech. Therefore, some elements of music (*e.g.,* tonality and meter) may assist them in producing structured and hierarchical sound sequences. The results may suggest a connection between musicality and the production of long-form speech, at least in the earliest stage of the evolution of vocal language in humans (Brown, 2017).

## Acknowledgements

This study was supported by MEXT/JSPS Grant-in-Aid for Scientific Research on Innovative Areas #4903 (Evolinguisitics) JP17H06380.

## References

Savage, P., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. (2021). Toward inclusive theories of the evolution of musicality. *Behavioral and Brain Sciences*, *44*, E121.

Baptista, L. F., & Keister, R. A. (2005). Why birdsong is sometimes like music. *Perspectives in Biology and Medicine*, *48*(3), 426-443.

Heinsohn, R., Zdenek, C. N., Cunningham, R. B., Endler, J. A., & Langmore, N. E. (2017). Tool-assisted rhythmic drumming in palm cockatoos shares key elements of human instrumental music. *Science Advances*, *3,* e1602399.

Le Covec, M., Aimé, C., & Bovet, D. (2019). Combinatory sound object play in cockatiels: a forerunner of music? *Behaviour, 156*(5-8), 595-617.

Brown, S. (2017). A joint prosodic origin of language and music. *Frontiers in Psychology*, *8,* 1894.

# INVESTIGATING CROSS-LINGUISTIC VARIABILITY IN PSYCHOLINGUISTIC DIFFICULTY THROUGH INFORMATION DENSITY VARIABILITY MEASURES

Arturs Semenuks*[*1,2]

[*]Corresponding Author: asemenuks@ucsd.edu
[1]Department of Cognitive Science, UCSD, La Jolla, USA
[2]Center for Academic Research and Training in Anthropogeny, UCSD, La Jolla, USA

Are all languages equally complex? In recent years, this question has attracted attention of researchers from multiple scientific fields (Miestamo, 2017; Ehret, Blumenthal-Dramé, Bentz, & Berdicevskis, 2021). If language is a complex adaptive system (Beckner et al., 2009), we expect languages to differ in their complexity, and a growing number of studies indeed provide support for that position (McWhorter, 2001; Lupyan & Dale, 2010; Trudgill, 2011; Bentz & Winter, 2014; Ehret et al., 2021). This leads to an important question from the perspective of evolutionary linguistics: what leads to the changes in complexity that languages exhibit?

One of the most widely discussed hypotheses attempting to explain (some aspects of) cross-linguistic variability in complexity suggests that languages adapt to the cognitive and sociocultural niches they inhabit, i.e. languages change in complexity to accommodate the cognitive and communicative constraints of their speakers, resulting in, for example, language simplification in communities with more speakers or a higher percentage of non-native learners (Lupyan & Dale, 2010; Bentz & Winter, 2014).

Given this, we can expect a relation between descriptive complexity and psycholinguistic difficulty – less complex languages should be easier to learn, produce and/or comprehend. However, this assumption still requires empirical support, as theory-based complexity and empirically observed cognitive difficulty do not necessarily entail one another (Miestamo, 2017). Surprisingly, the existence of this relation is underresearched. Furthermore, while some studies support this hypothesis (Berdicevskis & Semenuks, 2022), a number of recent studies do not find support for the theorized links between facets of complexity and psycholinguistic difficulty (Atkinson, Smith, & Kirby, 2018; Semenuks & Berdicevskis, 2018; Wagner, Smith, & Culbertson, 2019; Johnson, Gao, Smith, Rabagliati, & Culbertson, 2021). Thus, the question of whether descriptively simpler languages are also easier does not yet have a clear answer, which makes it less clear why the observed cross-linguistic variability in complexity and the relationships between different complexity dimensions and properties of language speakers and communities exist.

One way to tackle this issue is to use metrics that are more informed by psycholinguistic research and can be reasonably assumed to be transparently related to (or even operationalize) facets of learning, production or comprehension difficulty. A promising candidate for such a metric is the average information density variability of a language at a particular structural level. Information density is the amount of information transmitted per unit at a particular level of organization (e.g. word or syllable), and it can be operationalized as the information content of the unit in its context, i.e. the negative log probability of the unit given its context. For example, in the sentence “I like coffee with milk and sugar”, the word “like” has a relatively high information density, as it is relatively unexpected given the preceding context, is more surprising, and has a higher information content, whereas “sugar” has a relatively low information density due to its high predictability (low information content) at the end of the sentence. Theoretical considerations and empirical research suggest that the difficulty of processing a linguistic unit is predicted by its information density. Based on this, Jaeger and Levy (2006) put forward the Uniform Information Density (UID) hypothesis, which argues that speakers aim to minimize the variability in the information density of their utterances, as a UID strategy minimizes the total difficulty of processing an utterance. A variety of studies provide empitical support for the UID hypothesis, e.g. see Genzel and Charniak (2002), Aylett and Turk (2004), Frank and Jaeger (2008).

Thus, we can expect that languages argued to be under higher pressure to be more efficiently structured, such as languages with more speakers, should be more easily processed, and thus have more constrained information density variability of units at different levels of organization. I investigate this hypothesis by calculating the variability in the word information density at the syntactic level using the part-of-speech annotated data for a sample of languages from the Universal Dependencies corpora. Mixed-effects models show that languages vary on this measure and, surprisingly, exhibit higher values of information density variability with more speakers. However, we also find that the average information density level decreases with the number of speakers and is negatively correlated with the information density variability, thus lowering the average unexpectedness of a syntactic unit and creating a tradeoff. Taken together, the results suggest that while individual facets of language complexity correlate with extralinguistic sociocultural properties, they also sometimes trade off with each other, potentially providing a solution to why some previously studied facets of complexity do not correlate with psycholinguistic difficulty.

## References

Atkinson, M., Smith, K., & Kirby, S. (2018). Adult learning and language simplification. *Cognitive science*, *42*(8), 2818–2854.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*(1), 31–56.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., et al.. (2009). Language is a complex adaptive system: Position paper. *Language learning*, *59*, 1–26.

Bentz, C., & Winter, B. (2014). Languages with more second language learners tend to lose nominal case. In *Quantifying language dynamics* (pp. 96–124). Brill.

Berdicevskis, A., & Semenuks, A. (2022). Imperfect language learning reduces morphological overspecification: Experimental evidence. *PloS one*, *17*(1), e0262876.

Ehret, K., Blumenthal-Dramé, A., Bentz, C., & Berdicevskis, A. (2021). Meaning and measures: interpreting and evaluating complexity metrics. *Frontiers in Communication*, *6*, 61.

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30).

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th annual meeting of the association for computational linguistics* (pp. 199–206).

Jaeger, T., & Levy, R. (2006). Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems*, *19*.

Johnson, T., Gao, K., Smith, K., Rabagliati, H., & Culbertson, J. (2021). Investigating the effects of i-complexity and e-complexity on the learnability of morphological systems. *Journal of Language Modelling*, *9*(1), 97–150.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1), e8559.

McWhorter, J. H. (2001). The worlds simplest grammars are creole grammars.

Miestamo, M. (2017). Linguistic diversity and complexity. *Lingue e linguaggio*, *16*(2), 227–254.

Semenuks, A., & Berdicevskis, A. (2018). What makes a grammar difficult? experimental evidence. In *The evolution of language: Proceedings of the 12th international conference (evolangxii)*.

Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity.* Oxford University Press.

Wagner, S., Smith, K., & Culbertson, J. (2019). Acquiring agglutinating and fusional languages can be similarly difficult: Evidence from an adaptive tracking study. In *Cogsci* (pp. 3050–3056).

# INVESTIGATING THE RELATIONSHIP BETWEEN NUMBER OF SPEAKERS, I-COMPLEXITY, AND E-COMPLEXITY

Arturs Semenuks*[*1,2]

[*]Corresponding Author: asemenuks@ucsd.edu
[1]Department of Cognitive Science, UCSD, La Jolla, USA
[2]Center for Academic Research and Training in Anthropogeny, UCSD, La Jolla, USA

A number of theoretical proposals and computational models suggest that the sociocultural niche a language occupies affects the morphosyntactic complexity of that language, e.g. see McWhorter (2001), Wray and Grace (2007), Trudgill (2011), Dale and Lupyan (2012), Spike (2017). A common denominator of many such proposals is the focus on the difference between exoteric and esoteric societies, i.e. societies more (exoteric) and less (esoteric) open to outsiders. Wray and Grace (2007) propose arguably the most detailed psycholinguistic explanation for how esotericity promotes morphosyntactic complexity to date, suggesting that shared insider knowledge and implicit encoding leads to opaque, irregular, and more complex forms of language[1].

Recently, correlational cross-linguistic and experimental studies have provided empirical support to these theories, e.g. Lupyan and Dale (2010), Bentz and Winter (2014), Sinnemäki and Di Garbo (2018), Koplenig (2019), Kocab, Ziegler, and Snedeker (2019), Raviv, Meyer, and Lev-Ari (2019), Berdicevskis and Semenuks (2020, 2022). It is encouraging that researchers arrive at similar conclusions using approaches differing in methodological details, which could be taken as showing the robustness of the hypothesized relationship. At the same time, what morphological complexity is and what measures best capture it is not yet settled (Berdicevskis et al., 2018; Ehret, Blumenthal-Dramé, Bentz, & Berdicevskis, 2021). Additionally, how different dimensions of complexity interact with each other is far from resolved and what specific dimension(s) of complexity are affected by sociocultural niche is not yet clear. For example, Sinnemäki and Di Garbo (2018) find their verbal morphological complexity measure to be correlated with the total amount of speakers a language has and the percentage of L2 speakers in its population, but find no similar relationship for nominal complexity. Thus the aforementioned variability in methodological details leaves open

---

[1] In turn, these properties themselves make it harder for outsiders to understand the language, creating a feedback loop.

some important questions, including (i) are languages in exoteric niches simpler on all dimensions of complexity? and (ii) how are different dimensions of complexity related?

Here I explore whether languages of more exoteric societies (operationalized as the total number of speakers) tend to have morphological paradigms of lower i-complexity. Following Ackerman and Malouf (2013), we define i-complexity as the average conditional entropy between the word forms in a paradigm, which captures the average amount of information needed for a speaker to predict all inflectional forms of a new lemma. On the one hand, it should be expected that lower values of i-complexity facilitate learning and would be under stronger selective pressure in exoteric societies with more speakers. On the other hand, some information theoretic measures of language structure have been reported to be highly similar across languages due to psycholinguistic constraints (Coupé, Oh, Dediu, & Pellegrino, 2019). As Ackerman and Malouf (2013) hypothesize that i-complexity is constrained to facilitate preservation of efficient linguistic structures, we can similarly expect it to not differ substantially across languages.

Additionally, recent studies have provided evidence for a negative correlation between i-complexity and e-complexity, i.e. the number of morphosyntactic distinctions that a particular language makes (Cotterell, Kirov, Hulden, & Eisner, 2019; Johnson, Gao, Smith, Rabagliati, & Culbertson, 2021). I extend the previous analyses using methods more closely aligned with the proposal in Ackerman and Malouf (2013) and following the advice for best practices and directions in Malouf, Ackerman, and Semenuks (2020).

I use the data (71 languages, 47 language families) from the Surrey Morphological Complexity Database (Baerman, Brown, Evans, Corbett, & Cahill, 2015), as well as the data from the UniMorph 3.0 Project (60 languages, 11 language families) (McCarthy et al., 2020). The former database provides high quality annotated data for inflectional paradigms from a typologically and geographically diverse sample of languages, whereas the latter provides a less diverse sample, but allows for a more theory-neutral data-driven process of estimating paradigm information for a language. In the UniMorph 3.0 data, the set of noun paradigms for each language is estimated by removing the maximal shared subset of characters within word forms for all lemmas in a language. I use mixed-effect linear models to control for phylogeny and geography in analyzing the data. No significant correlation between i-complexity and the number of speakers a language is observed. However, similarly to Ackerman and Malouf (2013), languages are found to frequently have lower values than expected as shown through Monte Carlo simulations. Additionally, both data sets suggest a negative correlation between i-compexity and e-complexity. Taken together, the results suggest that i-complexity is not significantly affected by the sociocultural niche a language occupies, however we see evidence of optimization of this aspect of linguistic structure, potentially supporting its hypothesized psycholinguistic importance.

## References

Ackerman, F., & Malouf, R. (2013). Morphological organization: The low conditional entropy conjecture. *Language*, 429–464.

Baerman, M., Brown, D., Evans, R., Corbett, G. G., & Cahill, L. (2015). *Surrey morphological complexity database.* (University of Surrey. `http://dx.doi.org/10.15126/SMG.23/1`)

Bentz, C., & Winter, B. (2014). Languages with more second language learners tend to lose nominal case. In *Quantifying language dynamics* (pp. 96–124). Brill.

Berdicevskis, A., Çöltekin, Ç., Ehret, K., Prince, K. von, Ross, D., Thompson, B., Yan, C., Demberg, V., Lupyan, G., Rama, T., et al.. (2018). Using universal dependencies in cross-linguistic complexity research. In *Proceedings of the second workshop on universal dependencies (udw 2018).*

Berdicevskis, A., & Semenuks, A. (2020). Different trajectories of morphological overspecification and irregularity under imperfect language learning. *The Complexities of Morphology*, 283–305.

Berdicevskis, A., & Semenuks, A. (2022). Imperfect language learning reduces morphological overspecification: Experimental evidence. *PloS one*, *17*(1), e0262876.

Cotterell, R., Kirov, C., Hulden, M., & Eisner, J. (2019). On the complexity and typology of inflectional morphological systems. *Transactions of the Association for Computational Linguistics*, *7*, 327–342.

Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science advances*, *5*(9), eaaw2594.

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in complex systems*, *15*(03n04), 1150017.

Ehret, K., Blumenthal-Dramé, A., Bentz, C., & Berdicevskis, A. (2021). Meaning and measures: interpreting and evaluating complexity metrics. *Frontiers in Communication*, *6*, 61.

Johnson, T., Gao, K., Smith, K., Rabagliati, H., & Culbertson, J. (2021). Investigating the effects of i-complexity and e-complexity on the learnability of morphological systems. *Journal of Language Modelling*, *9*(1), 97–150.

Kocab, A., Ziegler, J., & Snedeker, J. (2019). It takes a village: The role of community size in linguistic regularization. *Cognitive Psychology*, *114*, 101227.

Koplenig, A. (2019). Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society open science*, *6*(2), 181274.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, *5*(1), e8559.

Malouf, R., Ackerman, F., & Semenuks, A. (2020). Lexical databases for computational analyses: A linguistic perspective. *Proceedings of the Society for Computation in Linguistics*, *3*(1), 297–307.

McCarthy, A. D., Kirov, C., Grella, M., Nidhi, A., Xia, P., Gorman, K., Vylomova, E., Mielke, S. J., Nicolai, G., Silfverberg, M., et al.. (2020). Unimorph 3.0: Universal morphology.

McWhorter, J. H. (2001). The worlds simplest grammars are creole grammars.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*(1907), 20191262.

Sinnemäki, K., & Di Garbo, F. (2018). Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in psychology*, *9*, 1141.

Spike, M. (2017). Population size, learning, and innovation determine linguistic complexity. In *Cogsci.*

Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity.* Oxford University Press.

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, *117*(3), 543–578.

# GRAMMATICAL COMPLEXITY IS ONLY WEAKLY INFLUENCED BY THE SOCIOLINGUISTIC ENVIRONMENT

OLENA SHCHERBAKOVA *1, SUSANNE MARIA MICHAELIS[2], HANNAH J. HAYNIE[3], SIMON J. GREENHILL[4], DAMIÁN E. BLASI [5, 2, 6], RUSSELL D. GRAY[2,7], VOLKER GAST[8], and HEDVIG SKIRGÅRD[2]

*Corresponding Author: shcherbakova@shh.mpg.de
[1]Max Planck Institute for the Science of Human History, Jena, Germany
[2]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[3]University of Colorado, Boulder, CO, USA
[4]School of Biological Sciences, University of Auckland, Auckland, New Zealand
[5]Harvard University, Cambridge, MA, USA
[6]Yale University, New Haven, CT, USA
[7]School of Psychology, University of Auckland, Auckland, New Zealand
[8]Friedrich-Schiller University of Jena, Jena, Germany

Recent studies claim that the social environment influences the evolution of language structures. In particular, grammatical complexity has been proposed to be lower in communities with looser social networks, higher numbers of L1 speakers, and higher proportions of L2 speakers (among others, Kusters 2003, Trudgill 2011, Lupyan & Dale 2010, Sinnemäki & Di Garbo 2018). The explanation for these relationships relies on the assumption that larger communities are exposed to more contact than smaller ones. Specifically, due to substantial proportions of L2 speakers in large communities, the more complex features are not always transmitted to further generations, and hence languages become simpler over time. Here we test these claims on a global scale using two metrics of grammatical complexity while controlling for phylogenetic and spatial non-independence.

In previous studies, metrics of grammatical complexity often capture two different phenomena (cf. Lupyan & Dale 2010): the amount of phonologically fused marking ("boundness") and the number of semantic distinctions ("informativity"). In our study, we separate these into two different metrics to evaluate potential differences in their behavior.

We use the typological information from Grambank (The Grambank Consortium 2022) to construct metrics for each of these concepts and test whether variation in the metric scores depends on the number of L1 speakers, the proportion of L2 speakers, the status of the language (official/not official) obtained from Ethnologue (Eberhard et al. 2021), number of linguistic neighbors, and population density available in Bromham et al. (2022). The status of the language has not been previously used in investigating these questions, but we include it in this study to glean the information on the function of the language in a community. In line with previous research, we expect the two dimensions of grammatical complexity, boundness and informativity, to be negatively correlated with all social variables, except for population density which serves as a proxy for social network density and can be expected to positively correlate with complexity dimensions (c.f. Trudgill 2011).

To explore relationships between language and social structures, we adopt a spatiophylogenetic modeling technique introduced in Dinnage et al. (2020), a Bayesian approach that uses an Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009, Martins et al. 2013). The models fit social variables as fixed effects and two structured random effects to estimate the phylogenetic and spatial influence on variation in metric scores. This method allows us to not only control for phylogenetic and spatial non-independence of languages but also to estimate the effects of phylogeny and geographical distance between languages on variation in grammatical complexity and evaluate if adding the social variables improves the explanatory power of the model above and beyond the phylogenetic and geographic similarity. We compare the models combining sets of random and fixed effects based on obtained WAIC values (Watanabe 2010).

Contrary to prior studies, our Bayesian spatiophylogenetic modeling results do not support a hypothesis that morphological complexity (boundness) is strongly influenced by the number of L1 speakers, the proportion of L2 speakers, the number of linguistic neighbors, or population density. In contrast to previous studies, we disentangle boundness from informativity in our analyses, and our findings can be explained by the use of an extensive dataset and our rigorous control for phylogenetic and spatial non-independence of languages. As a result, the claimed link between the grammatical complexity and social variables does not appear to be strong and is found only weakly between informativity scores and some of the social variables, such as a positive effect of L1 speaker population and the official language status on informativity. This indicates that the evolution of these complexity dimensions is better explained from the perspective of inheritance and areal diffusion rather than constraints imposed by different social environments.

## Acknowledgements

The authors are thankful to Sam Passmore for his advice on the methods.

## References

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, *6*(2), 163–173.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2020). *Ethnologue: Languages of the World* (24th ed.). SIL International. www.ethnologue.com

Dinnage, R., Skeels, A., & Cardillo, M. (2020). Spatiophylogenetic modelling of extinction risk reveals evolutionary distinctiveness and brief flowering period as threats in a hotspot plant genus. *Proceedings of the Royal Society B*, *287*(1926), 20192817.

Kusters, W. (2003). *Linguistic complexity*. Netherlands Graduate School of Linguistics.

Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure. *PLOS ONE*, *5*(1), e8559. https://doi.org/10.1371/journal.pone.0008559

Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, *67*, 68–83.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series b (Statistical Methodology)*, *71*(2), 319–392.

Sinnemäki, K., & Di Garbo, F. (2018). Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology*, *9*, 1141.

The Grambank Consortium (eds.). (2022). *Grambank 1.0.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://grambank.clld.org

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. OUP Oxford.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

# CONSTRAINTS ON COMMUNICATING ORDER OF EVENTS IN PANTOMIME

MARTA SIBIERSKA*[1,2], PRZEMYSŁAW ŻYWICZYŃSKI[1], JORDAN ZLATEV[2], MONIKA BORUTA-ŻYWICZYŃSKA[1], JOOST VAN DE WEIJER[3], ADAM GUTOWSKI[1], and MARIUSZ LEWANDOWSKI[1]

*Corresponding Author: sibier@umk.pl
[1]Centre for Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland
[2]Division for Cognitive Semiotics, Lund University, Sweden
[3]The Humanities Lab, Lund University, Sweden

Pantomime, understood as a self-sufficient, predominantly non-conventional means of communication based on bodily mimesis (Żywiczyński et al. 2018), has become a strong candidate for "the original human-specific communicative system" (Zlatev et al. 2020). Although pantomime affords successful communication in many contexts, it also has its limitations (see e.g. Żywiczyński et al. 2021). In this study, we looked at its constraints on communicating order of events. We assumed that as pantomime unfolds in real time, it is fit for conveying simple narratives, where events are arranged in a chronological sequence. At the same time, it is less suitable for expressing events in a non-chronological order, characteristic of more complex narratives (Boyd 2017; Żywiczyński et al. 2018). This assumption was put to test in a study based on basic "semiotic games" (see e.g. Krauss & Winheimer 1966; Fay et al. 2010; cf. Fay et al. 2013, 2014).

The study consisted of 4 games, during which 52 participants, working in pairs, took turns as directors and matchers. The director was presented with a verbal representation of a story consisting of three events. In the chronological condition, the story was arranged so that the resultative event was at the end (e.g. "A man opened the door. Then he saw a bear. Then he ran away"). In the non-chronological condition, the resultative event was placed at the beginning (e.g. "A man ran away, because when he opened the door, he saw a bear"). The task of the director was to communicate the story to the matcher by means of body movements. The matcher was presented with 4 comic strips: one which corresponded to the director’s input and three distractors. Their task was to choose the correct alternative. Communicative success was operationalised as the accuracy of the matchers’ responses.

We found that non-chronological representations had lower communicative success (34% correct matches) when compared to representations of chronological order of events (94% correct matches). To explore this result, we used a logistic regression model which included an interaction between two predictors (condition and game), and pairs of participants as a random effect. The model showed significant differences in communicative success between the conditions, as well as between the games (see Fig. 1).

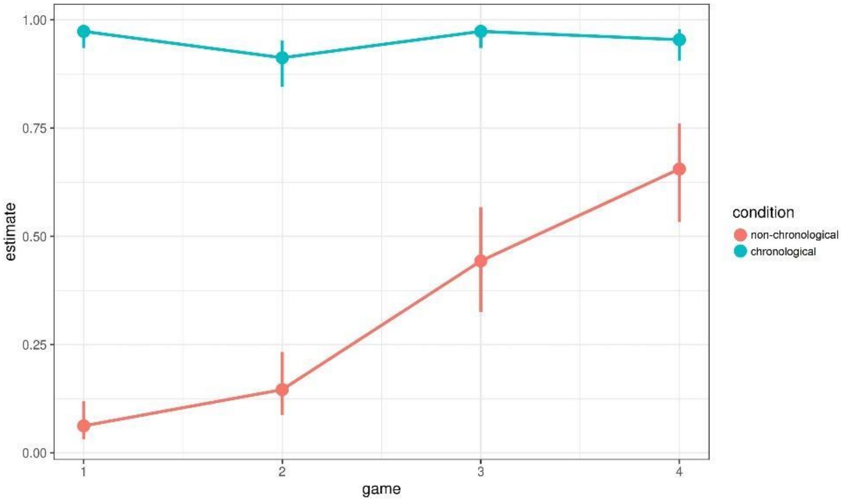


Figure 1. Predicted proportions of correct matches with 95% intervals, corresponding to the log odds from the logistic regression model described above. Post-hoc tests indicated that differences between all the games in the non-chronological condition were significant.

A possible explanation of this result is that some pairs of participants managed to indicate a disruption of chronology by means of specific movements, which conventionalised over the games. These results suggest that while it is fairly easy to communicate events in a chronological order in pantomime, the need to convey more chronologically complex narratives calls for a successful negotiation of specific communicative strategies over time, thus pushing pantomime towards “protolanguage”. We discuss these findings in the context of the bodily mimesis theory and the evolutionary trajectory of pantomime and narrative (e.g. Arbib 2012; Collins 2013; Donald 2001; Zlatev 2014; Zlatev et al. 2020).

## Acknowledgements

This research was funded by the Polish National Science Centre (NCN) under the agreement UMO-2017/27/B/HS2/00642. Marta Sibierska was funded by the Polish National Agency for Academic Exchange (NAWA) under the agreement PPN/BEK/2020/1/00319/U/00001.

## References

Arbib, M. A. (2012). *How the brain got language*. Oxford: Oxford University Press.

Boyd, B. (2017). The evolution of stories: From mimesis to language, from fact to fiction. *WIREs Cognitive Science*, 9(1), 1444, doi: 10.1002/wcs.1444.

Fay, N. et al. (2014). Creating a communication system from scratch: gesture beats vocalization hands down. *Frontiers in Psychology*, 5, 354, doi: 10.3389/fpsyg.2014.00354.

Fay, N., Arbib, M. A., Garrod, S., (2013). How to bootstrap a human communication system. *Cognitive Science*, 37, 1356–1367, doi: 10.1111/cogs.12048.

Fay, N. et al. (2010). The interactive evolution of human communication systems. *Cognitive Science*, 34(3), 351–386.

Krauss, R. M. & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4, 343–346.

Zlatev, J., Żywiczyński, P., Wacewicz, S. (2020). Pantomime as the original human-specific communicative system. *Journal of Language Evolution*, 5(2), 156–174, doi: 10.1093/jole/lzaa006.

Żywiczyński, P. et al. (2021). Evolution of conventional communication. A cross-cultural study of pantomimic re-enactments of transitive events. *Language & Communication*, 80, 191–203.

Żywiczyński, P., Wacewicz, S., Sibierska, M. (2018). Defining pantomime for language evolution research. *Topoi*, 37(2), 307–318.

# SHALL I POINT IT OUT?

REBECCA SLAGLE LUENSER* 1[1] and HEIDI LYN 2 [1,2]

*Corresponding Author: rl2122@jagmail.southalabama.edu
[1]Department of Psychology, University of South Alabama, Mobile AL USA
[2]Joan M. Sinnott Chair of Psychology and Department of Marine Sciences, University of South Alabama, Mobile AL USA

## 1. Ostensive Communication

A great debate exists when it comes to ostensive (intentional) communication and non-human primates. Some argue that non-human primates likely lack ostensive communication or are at least very limited in it (Scott-Phillips, 2015). Scott-Phillips (2015) contends that this limitation is due to great ape communication relying on a code model in contrast to human communication's ostensive-inference model. Still, others argue that the behaviors that suggest ostensive communication in human research exist in apes as well, including eye contact, and goal-directed, intentional communication (Moore, 2015).

Previous research has found bonobos in environments with rich language exposure are quite successful at following pointing gestures (Lyn et al., 2010). The bonobos studied here are unique in their high levels of exposure to natural English and lexigrams (Savage-Rumbaugh, 1993). The bonobos have also successfully passed previous pointing tests, likely due to daily pointing exposure. These environmental conditions gave Kanzi and the other bonobos a possible advantage in displaying their comprehension of ostensive communications.

Here, we separated the act of pointing into two components: the point mechanism (the gesture) and the ostensive communication (eye contact and gaze direction) in the hopes of exploring whether the bonobos understand pointing as a simple association, or as we hypothesize, they understand the communicative intent, including the ostensive cues that accompany the act itself.

## 2. Methodology

The present study used an object-choice task in which six bonobos were each tested on contralateral point gestures in five conditions in which the eye gaze and attention of the researcher were manipulated. Subjects were presented with two paper bags in front of their home enclosure, one which contained a small food reward. A researcher pointed to the baited bag contralaterally, focusing attention as described by each condition: (1) Contralateral (2) Drop Gaze, (3) Object-Ape, (4) Gesture-Object and (5) Gesture-Ape. For the Contralateral condition, the researcher's attention shifted equally between the ape, the object, and the gesture. For the Drop Gaze, attention focused on either the gesture or the floor. For the remaining conditions (Object-Ape, Gesture-Object, and Gesture-Ape), attention shifted equally between the two targets. The trial was successful when the ape pointed to the correct object. The trial was unsuccessful when the ape failed to point within 15 seconds or pointed to the incorrect object.

## 3. Discussion

Our results support previous findings that language enculturated bonobos have some cognitive capacity for understanding ostensive communication (Lyn et al., 2010). Aligning with the hypothesis, the bonobos with a more extensive history of language exposure performed better than those with less exposure (Kruskal-Wallis test, chi-square = 54.8, df = 5; $p < .001$). Mali, who has had limited language enculturation, scored much lower than the other apes on all conditions (Dwass-Steel-Critchlow-Fligner pairwise comparisons for each ape, For each $p < .001$). Additionally, in contrast to pilot data, apes performed equally well on the drop-gaze condition as the more ostensive conditions, though results were not significant (Kruskal-Wallis test, chi-square = 5.90, df = 4; $p = 0.207$). This discrepancy may be due to altered methodologies, underscoring the need for care when designing studies of this nature. While some argue that ostension cannot be separated from the communication act itself, that they are inherently the same thing (Heintz & Scott-Phillips, in press), it is our plan to continue attempting to piece apart these components in further studies.

## References

Heintz, C., & Scott-Phillips, T. (in press). Expression unleashed: The evolutionary and cognitive foundations of human communication. Behavioral and Brain Sciences.

Lyn, H., Russell, J.L., & Hopkins, W.D. (2010). The impact of environment on the comprehension of declarative communication in apes. Psychological Science, 21(3), 360-365.

Moore, R. (2015). Meaning and ostension in great ape gestural communication. Animal Cognition, 19, 223-231.

Savage-Rumbaugh, E.S., Murphy, J., Sevcik, R.A., Brakke, K.E., Williams, S.L., & Rumbaugh, D.M. (1993) Language comprehension in ape and child. Monographs of the Society for Research in Child Development, 58(3-4), 1-222.

Scott-Phillips, T. C. (2015). Nonhuman primate communication, pragmatics, and the origins of language. Current Anthropology, 56(1), 56-80.

# SIMULTANEITY AS AN EMERGENT PROPERTY OF SIGN LANGUAGES

ANITA SLONIMSKA* [1,2], ASLI OZYUREK [1,2,3], and OLGA CAPIRCI [4]

*Corresponding Author: Anita.Slonimska@mpi.nl
[1] Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
[2] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[3] Donders Centre for Cognition, Nijmegen, The Netherlands
[3] ISTC, National Research Council, Rome, Italy

One of the defining properties of natural languages is segmenting holistic representations into smaller meaning units that allow to combine them into larger meaning units, leading to compositionality (Kirby et al., 2008). This has been shown to constitute an emergent property of linguistic systems that evolved initially from holistic representations to accommodate the pressures of communicative efficiency during language use and language transmission to new learners (e.g., Kirby et al., 2008; Motamedi et al., 2019; Senghas et al., 2004). One of the evidence for this claim comes from sign language emergence research which has shown the emergence of segmentation out of holistic representations and linear sequencing of these meaning units (Senghas et al., 2004). However, in sign languages, due to the affordances of the visual modality to use multiple articulators and iconicity, meaning units can be organized not only linearly but also simultaneously. In the present study, we investigate if simultaneity, in addition to linearity, is an emergent property of sign languages by comparing the use of simultaneous constructions in LIS (Italian Sign Language) to that of silent gestures used by hearing Italian speakers.

Recent research has shown that LIS signers use simultaneous and iconic constructions (i.e., diagrammatic iconicity) as modality-specific properties to achieve communicative efficiency through clustering related meaning units closer together when they are asked to encode informatively rich events in an interactive task (Slonimska et al., 2020). The study showed that as the events to be communicated became more informationally dense (i.e., contained more semantic information units) LIS signers increased the use of simultaneous constructions as well as information density of these constructions. However, it remained to be explored whether such simultaneity might constitute a general affordance of communication in visual modality rather than a linguistic property that has evolved for greater communicative efficiency. Therefore, here we asked whether hearing participants with no knowledge of any sign language using only their gestures to communicate could recruit multiple articulators and iconicity to represent multiple elements of the event simultaneously to the same extent as signers.

In the present study, we conducted the same experiment as in Slonimska et al. (2020) with 23 Italian speakers with no knowledge of any sign language (12 females, M age = 26.04) using silent gesture to describe the events of varying information density in a director-matcher task. We coded whether movement segments (MS), i.e., segments based on gesture strokes (Kendon, 2004) used by gesturers contained simultaneity, where more than one articulator is used to represent distinct semantic information units. We also coded the information density (i.e., the number of simultaneously represented semantic information units) of these MS. We then compared frequencies and information density of MS with simultaneity in descriptions of silent gesturers and LIS signers (N = 23).

The findings showed that in comparison to signers, silent gesturers used MS with simultaneous constructions less frequently than signers ($\beta$ = -0.98, CI[-1.23; -0.73], SE = 0.13, $z$ = -7.65, $p$ <.001), even though their use increased in both groups as information density of the pictures they needed to describe increased (Fig.1). Furthermore, when silent gesturers used simultaneous MS they were significantly less informationally dense ($\beta$ = -0.21, CI[-0.26; -0.17], SE = 0.02, $z$= -9.16, $p$ <.001). Simultaneous constructions were also qualitatively different in the two groups. Gesturers relied on imagistic iconicity to represent simultaneous aspects of the event (e.g., to hold and caress) in separate movement segments (Fig. 2). In turn, signers relied on diagrammatic iconicity and brought various information units encoded on different articulators together (also relating to the two different referents) in their movement segments (MS 3, 4 and 5).

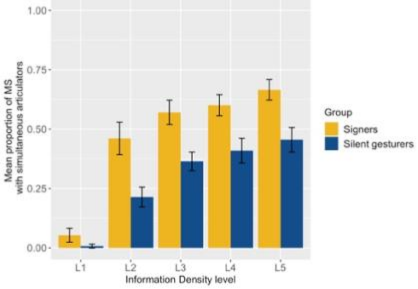


*Figure 1.* Mean proportions of simultaneous MS out of the total number of MS per trial for silent gesturers and signers when asked to describe events with increasing information density.

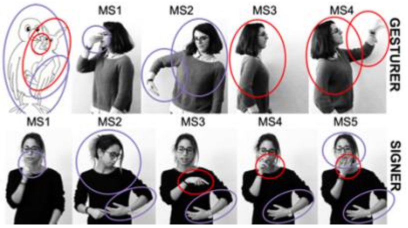


*Figure 2.* Prototypical encoding sequence for an event from silent gesturers and signers. Colored circles represent semantic information units of different referents (purple circles for the bird, red circles for the bunny).

The present study indicates that not only linear but also simultaneous expressions of segmented meaning units constitute an emergent property in sign languages that potentially evolved for achieving greater communicative efficiency. The findings highlight the role of modality-specific adaptive capabilities in linguistic expression and are relevant for the broader discussion about the multi-modal origins of language.

## References

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences, 105(27),* 9163-9168.

Kendon, A. (2004). Gesture: *Visible action as utterance.* Cambridge University Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105(31),* 10681-10686.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition, 192,* 103964.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua. *Science, 305(5691)*, 1779-1782.

Slonimska, A., Özyürek, A., & Capirci, O. (2020). The role of iconicity and simultaneity for efficient communication: The case of Italian Sign Language (LIS). *Cognition, 200,* 104246.

# THE ONTOGENY OF BEHAVIOURAL RESPONSES TO PANT HOOTS IN WILD CHIMPANZEES

ADRIAN SOLDATI[*1,2,3], PAWEL FEDUREK[3,4], GUILLAUME DEZECACHE[3,5], GERESOMU MUHUMUZA[3], KLAUS ZUBERBÜHLER[1,2,3], and JOSEP CALL[1]

[*]Corresponding Author: as391@st-andrews.ac.uk
[1]School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK
[2]School of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[3]Budongo Conservation Field Station, Masindi, Uganda
[4]University of Stirling, Faculty of Natural Sciences, Stirling, UK
[5]Université Clermont Auvergne, CNRS, LAPSCO, Clermont-Ferrand, France

## 1. Introduction

Comparative researchers study the vocal systems of animal species to identify the presence or absence of traits shared with human language to explain its origins (Fitch & Zuberbühler, 2013). Traditionally, the vocal development of non-human primates, like their adult vocal communication, has been regarded as rather inflexible and hard-wired, which led to the idea that few parallels exist with the highly flexible and socially mediated acquisition of human language (Egnor & Hauser, 2004; Owren et al., 2011). However, limitations to this view include being largely based on studies conducted on monkey species, on alarm calls, and in artificial or captive settings (Seyfarth & Cheney, 1997; Snowdon, 2009). The development of pant hoots, complex vocal sequences used by wild chimpanzees as long-distance social calls (Fedurek et al., 2014; Marler & Hobbett, 1975; Mitani & Nishida, 1993), is largely unknown. Pant hoots are the most frequently used calls by adults and are flexibly used across most behavioural contexts (Marler & Tenaza, 1977). Chimpanzees either spontaneously produce pant hoots or respond vocally to others' pant hoots, sometimes by chorusing together. Given that this call is very rarely produced by immature individuals, one approach to determine if and how social factors mediate its ontogeny is to investigate how young receivers respond when exposed to others' pant hoots.

## 2. Methods

We collected data in the Sonso community (*Pan troglodytes schweinfurthii*) of the Budongo Forest, Uganda, from six infants (0–4 years; M:F, 4:2), seven juveniles (5–9 years; M:F; 4:3), all dependent on their mother, as well as from 25 mature individuals (M:F, 13:12). We used focal animal sampling as the main method of data collection (Altmann, 1974). We recorded the behavioural activity of the focal, information about pant hoots produced by the focal and by other group members, behavioural changes of the focal after being exposed to others' pant hoots, and continuous party composition. We used long-term data collected by field assistants to assess the gregariousness of individuals by measuring how frequently they were observed in a party while not being focal followed.

## 3. Results

Immature individuals spontaneously called 10 times less often than mature individuals but responded vocally only 2 times less often. The two most common behavioural responses to pant hoots were a head movement towards the call and vocal responses. When compared to mature chimpanzees, immature individuals produced both responses less often. More specifically, the older offspring, the male offspring, and the offspring of more gregarious mothers were more likely to move their head towards calls. The older male offspring, the older offspring of more gregarious mothers, and the offspring of more gregarious mothers in the presence of adult males were more likely to respond vocally. In addition, immature individuals were more likely to respond vocally when their mother also produced a vocal response. Finally, we observed that the male offspring of more gregarious mothers were exposed to more pant hoots overall.

## 4. Discussion

Our study shows that the ontogeny of vocal responses undergoes developmental changes, varies according to sex, and is socially mediated. The rather slow ontogeny of pant hoots is consistent with the idea that vocal usage and comprehension learning are less hard-wired in social calls. Our observations are also in line with the idea that the development of social skills tends to occur earlier in male great apes, likely as a result of different selection pressures. Higher levels of social and communicative exposure likely increase opportunities for immature individuals to learn appropriate responses. Chorusing between mother and offspring might function as a bonding signal or as auditory reinforcement. While a clear distinction between the development of human language and that of primate vocalisations exists, we demonstrated that chimpanzees' vocal ontogeny can be flexible and mediated by social factors, contrary to previous assumptions.

## References

Altmann, J. (1974). Observational Study of Behavior: Sampling Methods. *Behaviour*, *49*(3/4), 227–267.

Egnor, S. E. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, *27*(11), 649–654.

Fedurek, P., Donnellan, E., & Slocombe, K. E. (2014). Social and ecological correlates of long-distance pant hoot calls in male chimpanzees. *Behavioral Ecology and Sociobiology*, *68*(8), 1345–1355.

Fitch, W. T., & Zuberbühler, K. (2013). Primate precursors to human language: Beyond discontinuity. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of Emotional Communication* (pp. 26–48). Oxford University Press.

Marler, P., & Hobbett, L. (1975). Individuality in a Long-Range Vocalization of Wild Chimpanzees. *Zeitschrift Für Tierpsychologie*, *38*(1), 97–109.

Marler, P., & Tenaza, R. (1977). Signalling behaviour of apes with special reference to vocalisations. In P. Marler, R. Tenaza, & T. A. Sebeok (Eds.), *How animals communicate* (pp. 965–1033). Indiana University Press, Bloomington.

Mitani, J. C., & Nishida, T. (1993). Contexts and social correlates of long-distance calling by male chimpanzees. *Animal Behaviour*, *45*(4), 735–746.

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, *73*(6), 530–544.

Seyfarth, R. M., & Cheney, D. L. (1997). Some general features of vocal development in nonhuman primates. In C. T. Snowdon & M. Hausberger (Eds.), *Social Influences on Vocal Development* (1st ed., pp. 249–273). Cambridge University Press.

Snowdon, C. T. (2009). Plasticity of Communication in Nonhuman Primates. In M. Naguib, K. Zuberbuhler, N. S. Clayton, & V. M. Janik (Eds.), *Advances in the Study of Behavior* (Vol. 40, pp. 239–276). Elsevier.

## Acknowledgments

We are grateful to the management, staff, field assistants, and researchers of the Budongo Conservation Field Station (BCFS) for their support and discussions. We thank the Uganda Wildlife Authority and the Uganda National Council for Science and Technology for permission to conduct the study. We are thankful to Charles Paxton for statistical advice. The study was funded by a St Leonard College Inter-University scholarship awarded to J.C. and K.Z., a Swiss National Science Foundation grant (310030_185324) awarded to K.Z., and individual mobility grants from Swissuniversities and St Leonard College

awarded to A.S.

# HOW SELF-DOMESTICATION AND PROSOCIALITY MAY SHAPE CROSS-MODAL LANGUAGE

KEES SOMMER*[1,2], CHRISTINE CUSKLEY[3], SIMON KIRBY[4] and TESSA VERHOEF[2]

*Corresponding Author: k.sommer@hhs.nl
[1]IT & Design, The Hague University of Applied Sciences, The Hague, Netherlands
[2]Leiden Institute for Advanced Computer Science, Leiden University, Leiden, Netherlands
[3]Center for Behaviour and Evolution, Newcastle University, Newcastle, UK
[4]Center for Language Evolution, University of Edinburgh, Edinburgh, UK

An emerging literature in language evolution has highlighted the key role of self-domestication, with at least two crucial features for language evolving in other species through domestication: vocal learning in birds and the recognition of communicative intent in dogs (Thomas & Kirby, 2018). Selection for less aggressive individuals is also associated with the appearance of higher prosociality, and closely linked to increased levels of serotonin and other 'bonding' neurochemicals (Hare, 2017). Interestingly, substances that interact with such neurochemicals can heighten integration of the senses (synesthesia) in humans (Brang & Ramachandran, 2008; Luke & Terhune, 2013). Both synesthesia and the related phenomenon of shared cross-modal mappings may have played a key role in the early evolution of language (Bankieris & Simner, 2015; Cuskley & Kirby, 2013; Imai & Kita, 2014) as a way to bootstrap shared linguistic form-meaning mappings. The current paper links self-domestication and cross-modality, using a task intended to enhance participants' prosociality and measuring their sensitivity to linguistic cross-modal associations.

A total of 62 participants were recruited at the University of Edinburgh and paid £5. Half of the participants performed a social clapping task with the experimenter using a composed rhythm, aimed at stimulating prosociality: synchronising through a rhythm has been shown to increase prosociality in individuals (Kirschner & Tomasello, 2010; von Zimmermann et al., 2018). The remaining participants completed an asocial clapping task, using a random rhythm with the computer. Participants in both groups then performed a task designed to test their cross-modal sensitivity to linguistic stimuli, where pre-recorded

Figure 1. Cross-modal interface. Left to right: Slider at initial position, at largest position and at smallest position.

pseudowords had to be matched with a shape that could be manipulated in size and weight through a circular slider (see figure 1). The pseudowords were designed based on previously found strong associative reactions between consonant voicing and visual weight, and vowel openness and visual size (Schmidtke, Conrad, & Jacobs, 2014; Cuskley, 2013), and were therefore distinctive for both these features, as well as place of articulation (bilabial vs alveolar) to create more variation in the data. This created a total of 8 pseudowords: /ipi/, /ibi/, /iti/, /idi/, /apa/, /aba/, /ata/, and /ada/. Lastly, all participants were asked to answer an open question, where word count was used as a prosociality measure (Baumsteiger & Siegel, 2019).

Since size and weight were simultaneously adjusted with one slider, we refer to size for the shape produced. Pseudowords were grouped into 4 types (ordered): 1: *closed-voiceless*, 2: *closed-voiced*, 3: *open-voiceless*, 4: *open-voiced*, since participants were expected to map closed vowels and voiceless consonants with smaller sizes, where openness was expected to have a stronger effect, and no strong relationship between size and place of articulation was expected. Size in relation to condition, word type and prosociality was analyzed using linear mixed-effects models with maximum likelihood estimation (P-values calculated using the Satterthwaite's method). Sizes produced for word types followed the expected cross-modal pattern (1<2<3<4). Shape sizes were significantly larger for *open-voiceless* than for *closed-voiceless* ($b = 0.97 \pm 0.28$ SEM, $P < 0.001$) and for *open-voiced* items relative to *closed-voiceless* items ($b = 1.02 \pm 0.28$ SEM, $P < 0.001$). This also interacted with task final measures of prosociality: participants with a higher prosociality score had even larger shapes for both *open-voiceless* ($b = 0.39 \pm 0.18$ SEM, $P = 0.03$) and *open-voiced* words ($b = 0.40 \pm 0.18$ SEM, $P = 0.03$) relative to *closed-voiceless* items, suggesting a link between prosociality and cross-modal associations. However, the effect of the clapping task is less clear: social clapping only affected *open-voiceless* items, actually dampening associations relative to asocial clapping ($b = -0.36 \pm 0.15$ SEM, $P = 0.02$). Differences between the social and asocial clapping tasks may not have been sufficient, since both involved rhythmic coordination. Since the clapping task did not measurably influence cross-modal sensitivity but prosociality did, perhaps the effect relates more strongly to more stable neurochemical interactions. For example, while mu-opioids are more sensitive to prosocial manipulations (Manninen et al., 2017), serotonin is much more stable (Mitchell, 2006). Future work in this direction could take this into account.

Overall, this experiment presents an initial exploration of a role for prosociality in mechanisms facilitating intuitive meaning sharing. This could potentially open up new comparative and experimental research directions in the context of self-domestication.

## References

Bankieris, K., & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition, 136*, 186-195.

Baumsteiger, R., & Siegel, J. T. (2019). Measuring prosociality: The development of a prosocial behavioral intentions scale. *Journal of personality assessment*, *101*(3), 305-314.

Brang, D., & Ramachandran, V. S. (2008). Psychopharmacology of synesthesia; the role of serotonin S2a receptor activation. *Medical hypotheses*, *70*(4), 903-904.

Cuskley, C. (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics*, 5(1), 39-62.

Cuskley, C. and Kirby, S. (2013). Synaesthesia, cross-modality and language evolution. In: Simner, J. and Hubbard E.M. (Eds) Oxford Handbook of Synaesthesia, pp. 869-907.

Hare, B. (2017). Survival of the friendliest: Homo sapiens evolved via selection for prosociality. *Annual review of psychology*, *68*, 155-186.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological sciences*, *369*(1651), 20130298.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, *31*(5), 354-364.

Launay, J., Tarr, B., & Dunbar, R. I. (2016). Synchrony as an adaptive mechanism for large-scale human social bonding. *Ethology*, *122*(10), 779-789.

Luke, D., & Terhune, D. B. (2013). The induction of synaesthesia with chemical agents: a systematic review. *Frontiers in psychology*, *4*, 753:1-12.

Manninen, S., Tuominen, L., Dunbar, R. I., Karjalainen, T., Hirvonen, J., Arponen, E. & Nummenmaa, L. (2017). Social laughter triggers endogenous opioid release in humans. *Journal of Neuroscience*, *37*( 25), 6125-6131.

Mitchell, A. J. (2006). Two-week delay in onset of action of antidepressants: new evidence. *The British Journal of Psychiatry*, *188*( 2), 105-106.

Schmidtke, D., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in psychology*,*5*,80.

Thomas, J., & Kirby, S. (2018). Self-domestication and the evolution of language. *Biology & philosophy*, *33*(1), 1-30.

von Zimmermann, J., Vicary, S., Sperling, M., Orgs, G., & Richardson, D. C. (2018). The choreography of group affiliation. *Topics in Cognitive Science*, *10*(1), 80-94.

# COMBINING POINTING AND LANGUAGE DURING THE EARLY STAGES OF DEVELOPMENT: A CASE STUDY OF RUSSIAN AND CHINTANG

Sabine Stoll[1,2], Jekaterina Mažara*[1,2], and Elena V. M. Lieven[3]

*Corresponding Author: jekaterina.mazara@uzh.ch
[1]Department of Comparative Language Science University of Zurich, Switzerland
[2]Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland
[3]Emeritus Professor, Division of Human Communication, Development and Hearing, University of Manchester, United Kingdom

Both language and pointing are universal features and central components of human societies (cf. Kita, 2003). A milestone of early ontogeny is learning how to communicate intentionally. One of the first entry points into this process is pointing (Bates, 1976; Iverson & Goldin-Meadow, 2005). A number of cross-cultural studies have shown that pointing emerges around the same age and is used for similar communicative functions cross-culturally (Butterworth, 2003; Butterworth & Morissette, 1996; Callaghan et al., 2011; Carpenter, Nagell, & Tomasello, 1998; Liszkowski, Brown, Callaghan, Takada, & Vos, 2012; Ohama, 1984).

Pointing alone, however, is often under-specified without additional linguistic information and adult language users generally accompany points with language (Greenfield & Smith, 1976; Enfield, 2009). While the co-development between gestures and language is well-studied (Iverson & Goldin-Meadow, 2005), so far, little research has been dedicated to the specific content of point-accompanying language. Here, we examine the relationship between pointing and accompanying vocalizations and utterances in the production of **1 to 4-year-old children** in two very different cultural settings, **Russia (St. Petersburg)** and **Chintang** (a rural subsistence community in Eastern Nepal). We analyze the speech acts children perform and the information they convey while pointing to assess whether the way the two domains interact is independent of language and culture. Even though the specific systems of language and pointing differ greatly between cultures (Wilkins, 2003), the development of this interplay is a candidate for a universal of human communication.

In this pilot study, we narrowed our focus on the development of utterances and vocalizations accompanying *finger points*, since they are proposed to be the most commonly and probably most frequently used type of pointing by young children across cultures (Butterworth, 2003). Even though index finger pointing is ubiquitous in the world's societies, pointing behavior is an integral part of the style of communication between parents and children and might, thus, be subject to specific cultural reinforcement that would result in a different path of development (Masataka, 2003). To understand how children and adults behave in natural communicative situations, we rely on observational longitudinal data. We evaluate whether pointing behavior differs between adults and children of the two cultures and how closely the children's production is linked to child-surrounding adult input. The video and audio data was hand-annotated for type of point as well as type of sentence, intent, and content of accompanying utterances/vocalizations.

First, we compared the proportion of finger points in both cultures in the target children's and adults' production. Second, we examined the development of accompanying utterances to understand how often pointing occurs without accompanying utterances, whether this changes with development, and whether it differs from the production in the ambient language. In a last step, we take a more detailed look at the content of the utterances that accompany finger pointing in both languages.

We found that finger points represent a larger proportion of all points in Russian (65%) than in Chintang (35%). Russian children tend to point more frequently but both Chintang and Russian children followed the distribution of point types found in their surrounding adults. Both Chintang and Russian adults use significantly fewer points unaccompanied by language or vocalizations than the children and neither children nor adults differ significantly between cultures. Over time, fewer of the children's points are silent or accompanied by vocalizations, which reflects their linguistic maturation. This is also reflected in the changes of the utterances' content in children's production over time. Children in the older age groups show a diversification of the information content they try to convey while pointing. 1 to 2-year-olds predominantly use finger points and vocalizations to draw their interlocutor's attention, older children increasingly point while making an additional comment about the referent they are pointing out. They also diversify the sentence types they use (declaratives, interrogatives, imperatives).

The composition of information content and sentence types in the input changes less across all recordings but, over time, we saw a reduction in simple identifications by adults. Across both languages and age groups (excluding the youngest children's production), the most commonly used utterances occurring with finger points are statements. Points and language are used in symbiosis, each adding information to the other.

This case study contributes to understanding the role of language in communicative pointing during the earliest stages of language development. Children in both languages differ from the input in term of content of their utterances but display cross-cultural similarities. Despite the very different cultural environments, overall gestural behavior, and adults' attitudes towards linguistic instruction, the interaction of finger pointing with language shows considerable similarities.

## References

Bates, E. (1976). *Language and context: The acquisition of pragmatics.* Academic Press.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where language, culture and cognition meet* (pp. 9–33). New York: Erlbaum Associates.

Butterworth, G., & Morissette, P. (1996). Onset of pointing and the acquisition of language in infancy. *Journal of Reproductive and Infant Psychology*, *14*(3), 219–231.

Callaghan, T. C., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., & Tomasello, M. (2011). *Early social cognition in three cultural contexts.* Wiley-Blackwell.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i–143.

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances* (Vol. 8). Cambridge University Press.

Greenfield, P. M., & Smith, J. H. (1976). *The structure of communication in early language development.* New York, NY: Academic Press.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371.

Kita, S. (2003). Pointing: A foundational building block of human communication. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 1–8). Mahwah, NJ: Lawrence Erlbaum.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & Vos, C. de. (2012). A prelinguistic gestural universal of human communication. *Cognitive science*, *36*(4), 698–713.

Masataka, N. (2003). From index-finger extension to index-finger pointing: Ontogenesis of pointing in preverbal infants. In S. Kita (Ed.), *Pointing: Where language, culture and cognition meet* (p. 69-84). Mahwah, NJ: Lawrence Erlbaum.

Ohama, K. (1984). Development of pointing behavior in infants and mother's responsive behavior: Longitudinal study of infants from 9 to 30 months. *The development of verbal behavior VI. Bulletin of the Faculty of Education, University of Tokyo*.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing: Where language, culture and cognition meet* (pp. 171–215). Erlbaum Mahwah, NJ.

# VERBAL LABELS INCREASE CONCEPTUAL ALIGNMENT

ELLISE SUFFILL[1,], JEROEN VAN PARIDON[1], and GARY LUPYAN*[1]

*Corresponding Author: lupyan@wisc.edu
[1]Dept of Psychology. University of Wisconsin-Madison, USA

Successful linguistic communication requires conversants to mean at least approximately the same thing by the same words. But how is this alignment achieved? One possibility is that participants have pre-existing concepts to which verbal labels are mapped. Alignment is then a matter of ensuring that in members of the same speech community, the same word points to the same concept. But how do the underlying conceptual representations become aligned in the first place? One source of alignment is shared sensory experiences mediated by similar perceptual systems. But is this enough? We test the possibility that language itself serves to align conceptual representations. Participants were asked to sort novel shapes and we measured the similarity between people's sorts. By separately manipulating previous perceptual experience with the shapes, and exposure to (entirely redundant) category labels, we tested (1) the role of shared perceptual experience and (2) the effect of labels on representational alignment. The results showed that shared experience with labels increased representational alignment more than shared perceptual experience alone. We consider the implications of this finding for the cognitive functions of language and for how language may be used to enable coordination in the face of non-shared perceptual experiences.

## Introduction

The idea that it is possible to communicate through direct transfer of mental states between minds has been a popular trope in science fiction for over a century. It has now has become the subject of empirical investigation, e.g., Rao et al. ask "Can information that is available in the brain be transferred directly in the form of the neural code, bypassing language altogether?" (2014). A common thread in science fiction treatments of telepathy and its modern revival is that natural language obscures communication because it is ambiguous, imprecise, and slow. For example, the computer scientist Yan LeCun, the chief AI scientist at Facebook, recently asserted that "Language is an imperfect, incomplete, and low-bandwidth serialization protocol for the internal data structures we call thoughts" (LeCun, 2021). Bypassing language is hence seen as a way of improving, or at least speeding up, communication.

The work we describe here is motivated by challenging a core assumption behind the telepathy trope: that our mental states are naturally aligned such that one person's thought is syntactically and semantically homologous to another person's. In the absence of this prior such alignment, transferring neural patterns that constitute mental states between people—even if technologically possible—would not lead to successful communication. What is at stake is important to the study of language evolution because understanding the evolution of a trait is greatly helped by understanding the range of its functions (Griffiths, 1993) and work on the functions of language are curiously under-represented in the study of language evolution (cf. Dessalles, 2007).

### *Conceptual alignment and language: what is the connection?*

Everyday communication seems to require conceptual alignment (e.g., Pickering & Garrod, 2021). When one person says "Pass the salt please" and another person passes them the salt, the two have achieved some amount of alignment: one person's request was successfully represented by the other. But where does this alignment comes from and what role, if any, natural language plays in establishing (rather than obscuring) it.

The idea that linguistic communication is possible only because our thoughts are already sufficiently aligned is a basic premise of philosophical positions such as Fodor's language of thought (Fodor, 1975). It is also a common starting point in theories of language learning that view children as mapping words onto pre-existing (and largely shared) concepts (Bloom, 2002; Pinker, 1994; Snedeker & Gleitman, 2004).

But another possibility is that alignment is achieved—in part—*through* language itself (e.g., Casasanto & Lupyan, 2014; Dingemanse, 2017; Gomila, 2011; Lupyan & Bergen, 2016). On this view, learning and using the syntax and semantics of a natural language helps people to structure and convey their thoughts in ways that are (more or less) understandable to others. Rather than just a device for conveying our thoughts, language provides an interface between minds (e.g., Clark, 1998; Gentner & Goldin-Meadow, 2003; Lupyan, 2012). If true, then telepathy could in principle be possible by transferring neural representations of words from one person to another, but such a scheme would not be the language-bypassing telepathy we've been promised, but rather an over-engineered form of texting.

The idea that language may play a causal role in promoting conceptual alignment is supported by several lines of evidence, some circumstantial, others more direct. First, there is the simple fact of substantial cross-linguistic diversity in all aspects of language (Evans & Levinson, 2009). If our conceptual representations were naturally aligned—either due to our shared biology, shared environment, shared goals, or all three—one might expect lexicons to show more similarity than they do. And although it is clear that the lexical systems of natural languages occupy a small space of all possible systems (e.g., Zaslavsky et al., 2018), it is striking that finding universal basic units of linguistic meaning has been so difficult. Even in the domain of perception, where one might find vocabularies to be most constrained by shared biology, one finds tremendous diversity of naming schemes (Majid, 2020; Majid et al., 2018). Diversity within a language is smaller (Forder & Lupyan, 2017) although it can depend on the measure one uses (Kuehni, 2004). Second, experimental evidence suggests that verbal labels can increase conceptual alignment across people, in both communicative (e.g., Markman & Makin, 1998) and non-communicative contexts (Suffill et al., 2016, 2019).

### *Current study*

Here we tested a strong version of the prediction that language promotes conceptual alignment. We exposed people to novel shapes grouped into two distinct categories (Figure 1). We then computed alignment among participants assigned to each of group using a sorting task. Our design allowed us to compare how conceptual alignment is affected by shared labels compared to alignment achieved through shaped perceptual experiences. Using materials with a clear pre-existing category structure allowed us to test whether labels help to align categories even when the existence of the categories is made plain by perceptual discontinuities (Fig. 1). This makes the current experiment substantially different from work examining the ways that labels can help mark distinctions in perceptually equidistant continua such as colors (Davidoff, 2001) and shapes (Plunkett et al., 2008) as well as from past work showing that labels facilitate the learning of new categories (Lupyan et al., 2007).

## Methods

### *Participants*

We recruited 129 (85 female, ages 18-22) psychology students from University of Wisconsin-Madison. Participants were randomly assigned to a *Baseline* (N = 45), *No Labels* (N = 43) or *With Labels* (N = 41) condition.

### *Materials*

We constructed two visual family-resemblance categories designed to be easy to distinguish and difficult to name. We began by generating two prototype shapes by creating a random collection of points and connecting them with a spline (Fig 1A, 1B). We then generated 18 exemplars per category by perturbing the points and fitting a new spline, creating low, medium, or high distortions (e.g., Fig. 1 bottom).

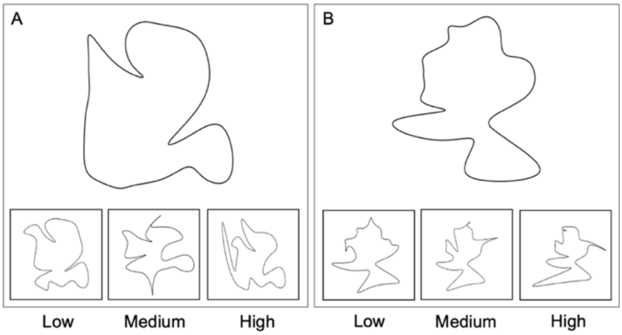


Figure 1. Category A (left) and B (right) prototypes with examples of "low", "medium" and "high" category exemplars.

### *Procedure.*

***Pre-exposure.*** Participants assigned to the *With Labels* or *No Labels* conditions began with a match-to-sample task designed to familiarize participants with the stimuli and have them repeatedly contrast within-category and between-category stimuli. On each trial, participants saw one of the shapes (the *standard*) for 1 sec followed by a 1 sec blank screen. Two shapes then appeared side by side. One of these was identical to the standard (*target*) and the other was a *foil*—a shape from the contrasting category. Participants had to choose which of the two shapes exactly matched the standard, i.e., on each trial they had to choose the target and not the foil. In the *With Labels* condition, the standard was presented with its corresponding nonsense label; half the *With Labels* participants heard category *A* shapes labeled as "a talp"; half heard them labeled as "a gek". Notice that the label is completely unnecessary for making a correct response and is therefore informationally redundant. Errors were signaled with a short buzzing sound. Participants assigned to the *Baseline* condition did not complete this phase and proceeded directly to the free sort.

***Free sort.*** We quantified alignment by measuring how participants in the three conditions arranged the shapes in a free sort—a common method for assessing people's conceptual representations (Goldstone, 1994; Kriegeskorte et al., 2008; Malt et al., 1999). Participants were shown 20 shapes (10 A shapes and 10 B shapes) arranged around the perimeter of the screen. These included 3 previously seen exemplars, 6 novel exemplars, and the previously unseen prototype. Participants were asked to cluster the shapes together in way that made sense to them, creating as many or as few clusters as they needed.

### *Analytic approach*

We computed alignment between people's item arrangements as follows: For each participant, we take the pairwise distances between all item pairs (20*19/2 = 190). We then compute the rank correlations between that participant's pairwise item distances and the pairwise item distances of the other participants in the same condition. The Fisher's z-transformed mean of these correlations represents the participant's average alignment to other participants. These are the values shown in Fig 2A. To statistically compare the groups in an unbiased way we counted each participant pairing as a single observation, but attributed the variance associated with this observation to both participants in the pairing using *lmerMultiMember* (van Paridon et al., 2022), an R package that allows for specifying multiple membership random effects. In addition, we computed for each participant a measure of *categoricality*, the median Euclidean distance between exemplars from different categories (e.g., $A_1$ and $B_2$) minus the median distance between exemplars from the same category (e.g., $A_1$ and $A_2$).

## Results

**Pre-exposure.** Accuracy on the delayed match-to-sample task was nearly identical for the *No Labels* ($M = 0.98$) and *With Labels* groups ($M= 0.98$). Given the task's simplicity, this was expected, and confirms that the categories were trivially easy to distinguish, regardless of labels.

**Categoricality.** Participants in all groups grouped within-category items closer together than between-category items: $\text{Categoricality}_{\text{baseline}}=151$ pixels, $\text{Categoricality}_{\text{no-labels}}=171$ pixels, $\text{Categoricality}_{\text{with-labels}}=264$ pixels. All three values were significantly greater than 0, $t$'s> 5, $p$<.0001, confirming that even

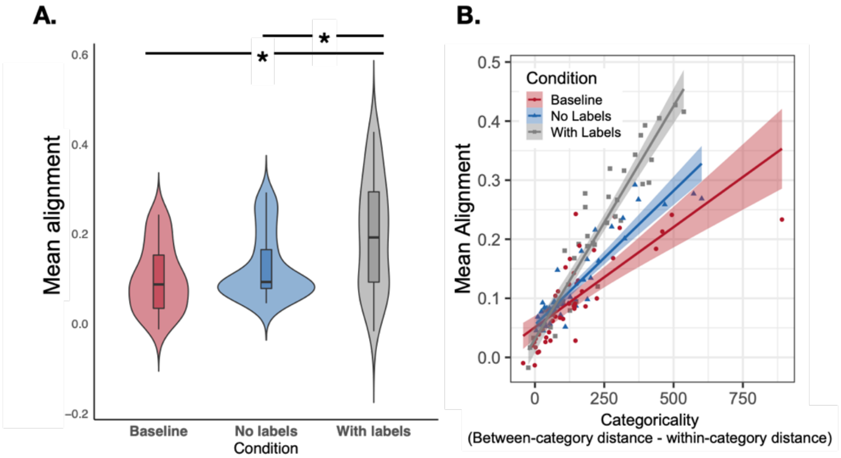


Figure 2. A. Mean alignment for the three tested conditions. B. The relationship between alignment and categoricality for each condition.

*Baseline* participants—who had no prior experience with seeing or contrasting the shapes—were still sensitive to the designed category structure. Categoricality was not significantly different between the *Baseline* and *No Labels* groups, $t<1$. In contrast, participants exposed to the nonsense labels produced more categorical sorts than those who had identical experience seeing and contrasting the shapes, but without being exposed to their names ($b$=93, $t$=2.4, $p$=.02).

**Alignment.** Average alignment for each group is shown in Fig. 2A. Those in the *Baseline* condition were as similar to one another in their shape arrangements as those who encountered the shapes several hundred times, but without the accompanying labels ($b$=.02, $t$=1.1, $p$=.24). In contrast, participants who were exposed to category labels, sorted shapes more similarly to one another than those in the *Baseline* condition ($b$=.09, $t$=4.9, $p$<.0001) and those in the *No Labels* condition ($b$=.07, t=3.7, p=.0004). The increase in alignment caused by labels was significantly greater than that caused by shared perceptual experiences ($t$=2.27, $p$=.03).

**Relationship between categoricality and alignment.** Why did labels increase alignment? One possibility is that alignment was mediated by categoricality: labels increased categoricality—leading to an increase in between category distance and a decrease in within-category, and these more tightly-clustered sorts were more aligned. As shown in Fig. 2B, there was indeed a strong relationship between categoricality and alignment (overall $r$=.83, p<.0001). And yet, categoricality (b=.08, t=21.9, p<.0001) and condition (b=.02, t=6.4, p<.0001) accounted for unique variance in predicting alignment; including categoricality made the effect of condition on alignment even stronger with the two predictors accounting for 83% of the variance. Fig. 2B also makes clear that there was a condition-by-categoricality interaction ($t$=7.71, $p$<.0001). For the same level of categoricality, exposure to labels yielded greater alignment. In short, labels increased both alignment and categoricality, but there was no evidence of mediation of alignment by categoricality.

## General Discussion

Language allows one to activate thoughts, old and new, in other people. The promise of telepathy—a direct exchange of mental states that bypasses natural language—is predicated on the assumption that thoughts are entirely independent of language; language is merely a medium by which the thoughts are transmitted. This assumption, however, may be wrong. The study we describe here provides a very limited, but nevertheless strong test of the hypothesis that even very stripped-down forms of language—redundant and seemingly uninformative verbal category labels—can increase conceptual alignment and do so to a greater extent than shared perceptual experiences alone. Our finding that verbal labels increase conceptual alignment is just one result using specific stimuli and task. Our hope is that future investigations can map out the generality of this result and the mechanisms by which labels achieve this effect.

The technology to transfer mental states may one day exist. Will it enable telepathy? The present results offer an early hint that however "imperfect and incomplete" language may be, attempts to bypass it may lead to a semantic disconnect and communicative failure. It may be possible to devise a system for re-aligning our thoughts into a mutually understandable form. Natural language is *just such a system* and the role natural language plays in aligning our thoughts may be another piece in the puzzle of its evolution.

## Acknowledgements

This work was supported by NSF-PAC Awards 2020969 to G.L.

## References

Bloom, P. (2002). *How Children Learn the Meanings of Words*. MIT Press.

Casasanto, D., & Lupyan, G. (2014). All Concepts are Ad Hoc Concepts. In E. Margolis & S. Laurence (Eds.), *Concepts: New Directions* (pp. 543–566). MIT Press.

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and Thought: Interdisciplinary themes* (pp. 162–183). Cambridge University Press.

Davidoff, J. (2001). Language and perceptual categorization. *Trends in Cognitive Sciences*, *5*(9), 382–387.

Dessalles, J.-L. (2007). *Why We Talk: The Evolutionary Origins of Language* (J. Grieve, Trans.). Oxford University Press, USA.

Dingemanse, M. (2017). On Brain-to-Brain Interfaces, Distributed Agency and Language. In N. J. Enfield & P. Kockelman (Eds.), *Distributed Agency* (pp. 59–66). Oxford University Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(05), 429.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard university press.

Forder, L., & Lupyan, G. (2017). Facilitation of color discrimination by verbal and visual cues. *Meeting of the Vision Sciences Society*.

Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Advances in the study of language and thought*. MIT Press.

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*(4), 381–386.

Gomila, T. (2011). *Verbal Minds: Language and the Architecture of Cognition* (1 edition). Elsevier.

Griffiths, P. E. (1993). Functional Analysis and Proper Functions. *The British Journal for the Philosophy of Science*, *44*(3), 409–422.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational Similarity Analysis—Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience*, *2*.

Kuehni, R. G. (2004). Variability in Unique Hue Selection: A Surprising Phenomenon. *Color Research & Application*, *29*(2), 158–162. https://doi.org/10.1002/col.10237

LeCun, Y. (2021, March 6). Language is an imperfect, incomplete, and low-bandwidth serialization protocol for the internal data structures we call thoughts. [Tweet]. *@ylecun*. https://twitter.com/ylecun/status/1368208931265388554

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Cognition*, *3*(54). https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G., & Bergen, B. (2016). How Language Programs the Mind. *Topics in Cognitive Science*, *8*(2), 408–424. https://doi.org/10.1111/tops.12155

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1082.

Majid, A. (2020). Human Olfaction at the Intersection of Language, Culture, and Biology. *Trends in Cognitive Sciences*.

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., Woll, B., LeLan, … Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369–11376.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*(2), 230–262.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, *127*(4), 331.

Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press.

Pinker, S. (1994). *The language instinct. William Morrow & Co*.

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681.

Rao, R. P. N., Stocco, A., Bryan, M., Sarma, D., Youngquist, T. M., Wu, J., & Prat, C. S. (2014). A Direct Brain-to-Brain Interface in Humans. *PLoS ONE*, *9*(11).

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. *Weaving a Lexicon*, *257294*.

Suffill, E., Branigan, H., & Pickering, M. (2016). *When the Words Don't Matter: Arbitrary labels improve categorical alignment through the anchoring of categories*. Proceedings of the 38th Meeting of the Cognitive Science Society, 1715–1720.

Suffill, E., Branigan, H., & Pickering, M. (2019). Novel labels increase category coherence, but only when people have the goal to coordinate. *Cognitive Science*, *43*(11), e12796.

van Paridon, J., Bolker, B., & Alday, P. (2022). *lmerMultiMember: Multiple membership random effects for mixed effects models in lme4* [R]. https://github.com/jvparidon/lmerMultiMember

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942.

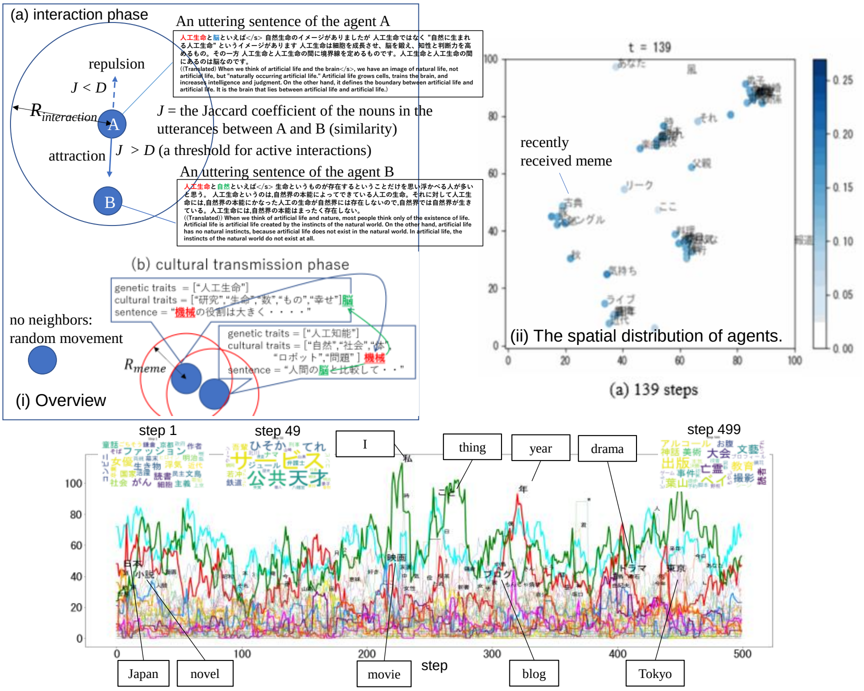
# A COMPUTATIONAL FRAMEWORK FOR STUDYING THE EVOLUTION OF CULTURAL TRAITS AMONG CHATTING AGENTS USING A GENERATIVE MODEL

Reiji Suzuki*, Soichiro Hirata and Takaya Arita

*Corresponding Author: reiji@nagoya-u.jp
Nagoya University, Japan

Constructive approaches have played a significant role in the understanding of cultural evolution and meme transmission, showing social relationships among hosts can affect the diversification of memes (Kobayashi et al. 2013, Weng et al. 2021). However, it is still unclear how novel memes emerge from complex communications in human language in the real world, such as chatting in a face-to-face situation and text-based communication in SNSs, and propagate in the population while a bigdata analysis contributed to clarifying the open-ended evolutionary dynamics of Web services such as social tags (Ikegami et al. 2019).

We propose a computational framework for studying the cultural evolution of memes among chatting agents using a generative model (Fig. 1 (i)). Individuals are represented as agents and arranged in a two-dimensional social space. The space abstracts their social relationships (i.e., the closer distance between agents represents more intimate relationships among them), which is inspired by the social particle swarm model (Nishimoto et al. 2013). When another neighboring agent is in its interaction radius, a communicative interaction occurs between them ((i)-(a)). Each agent has genetic and cultural traits, each composed of a list of words. The former represents non-evolvable and intrinsic preferences to the topic of a sentence to utter (e.g., gossip, (Mesoudi et al. 2006)). The latter represents the cultural ones, which evolve in the model. An agent utters a sentence produced by a pre-trained model of Japanese sentence generation (rinna/japanese-gpt2-small) based on GPT-2 (Radford et al. 2019). This model predicts and generates Japanese words that follow a given phrase. We generate an utterance of an agent by using a list of words in its genetic traits followed by its cultural traits as a given phrase, which is regenerated with a certain probability at each time step. We assume that agents who share more words among their utterances are



An uttering sentence of the agent A

人工生命と脳といえば</s> 自然生命のイメージがありましたが 人工生命ではなく "自然に生まれる人工生命" というイメージがあります 人工生命は細胞を成長させ、脳を鍛え、知性と判断力を高めるもの。その一方 人工生命と人工生命の間に境界線を定めるものです。人工生命と人工生命の間にあるのは脳なのです。

((Translated) When we think of artificial life and the brain</s>, we have an image of natural life, not artificial life, but "naturally occurring artificial life." Artificial life grows cells, trains the brain, and increases intelligence and judgment. On the other hand, it defines the boundary between artificial life and artificial life. It is the brain that lies between artificial life and artificial life.)

An uttering sentence of the agent B

人工生命と自然といえば</s> 生命というものが存在するということだけを思い浮かべる人が多いと思う。 人工生命というのは,自然界の本能によってできている人工の生命。それに対して人工生命には,自然界の本能にかなった人工の生命が自然界には存在しないので,自然界では自然界が生きている。人工生命には,自然界の本能はまったく存在しない。

((Translated)) When we think of artificial life and nature, most people think only of the existence of life. Artificial life is artificial life created by the instincts of the natural world. On the other hand, artificial life has no natural instincts, because artificial life does not exist in the natural world. In artificial life, the instincts of the natural world do not exist at all.

genetic traits = ["人工生命"]
cultural traits = ["研究","生命","数","もの","幸せ"] 脳
sentence = "機械の役割は大きく・・・・・"

genetic traits = ["人工知能"]
cultural traits = ["自然","社会","体", "ロボット","問題"] 機械
sentence = "人間の脳と比較して・・"

(iii) The change in the frequency of major uttered nouns and word clouds of unique nouns in several time steps.

Fig. 1: (i) Overview of the model and (ii, iii) results.

speaking about the topics more attractive to both. We measure the Jaccard index between the sets of the agents' uttering nouns. If the index is above the threshold value, the interparticle attraction occurs, and the repulsion occurs otherwise, of which strength is inversely proportional to the distance between the agents. We also introduce cultural transmission of traits into the interaction process ((i)-(b)). Each agent randomly extracts a noun from the sentence uttered by each closely neighboring agent with a certain probability and adds it to its cultural traits (and removes the oldest one if the number of traits exceeds the limit).

The preliminary experiments with agents, each of whose genetic traits had one of the names of famous novelists mainly in Heian - Reiwa periods in Japan, showed that agents were actively moving and exchanging novel traits with neighboring agents and the emergence and collapse of chatting clusters happened repeatedly (ii). In addition, the agents initially tended to utter words related to Japanese literature, and then these were taken over by more various words, and further trends of Japanese literature emerged (iii). These indicate that the proposed framework enables us to discuss the cultural evolution of novel traits in communicating agents in language.

## References

Ikegami, T., Hashimoto, Y. and Oka, M. (2019). Open-ended evolution and a mechanism of novelties in web services. *Artificial Life*, 25(2), 168-177.

Kobayashi, K., Suzuki, R. and Arita, T. (2012). Emergence of autocatalytic reaction in a meme propagation model based on particle motion. *Artificial Life and Robotics*, 17, 1, 92-96.

Mesoudi, A, Whiten, A. and Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology*, 97(3), 405-423.

Nishimoto, K., Suzuki, R. and Arita, T. (2013). Social particle swarm: explosive particle dynamics on cooperative/defective forces. In *Proceedings of IEEE ALIFE 2013 (IEEE SSCI 2013)* (pp. 134-139).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.

Weng, L. and Flammini, A., Vespignani, A. and Menczer, F. (2021). Blockchain and multi-agent system for meme discovery and prediction in social network. *Knowledge-Based Systems*, 229, 107368.

# AUDIO-VOCAL MIRROR NEURON IN THE SONGBIRD BASAL GANGLIA

YUKA SUZUKI*[1,2], SHIN YANAGIHARA[2], and KAZUO OKANOYA[2,1]

*Corresponding Author: suzuki-yuka236@g.ecc.u-tokyo.ac.jp
[1] Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
[2]Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan

Speech acquisition in humans is parallel to vocal learning in songbirds (Doupe & Kuhl, 1999; Prather et al., 2017). Both learnings are held by imitating adults' vocalizations and have critical periods that need auditory and motor information (Fromkin et al., 1974; Marler, 1970). In studying vocal learning of humans and songbirds, it is important to understand the neural mechanisms how to integrate auditory and vocal-motor information. Since songbirds have anatomically well-defined brain circuits dedicated to vocal learning, we focused on songbirds to understand the neural mechanisms underlying the audio-vocal integration at the cellular levels that is hard to research on human subjects.

The cortico-basal ganglia circuitry, which includes premotor cortical nucleus (HVC) and a part of the basal ganglia (Area X), is crucial for vocal learning in a songbird (Kao et al., 2005). HVCx neurons, which project to Area X, fire both when the bird is singing and when it is listening to the bird's own song (BOS) (Prather et al., 2008; Hessler & Okanoya, 2018). These neurons are called "audio-vocal mirror neurons". Audio-vocal mirror neurons should have important roles to integrate auditory and vocal motor information. Because HVCx neurons projects to Area X that is important for vocal leaning, there should be audio-vocal mirror neurons not only in HVC but also in Area X. In this study, we used Bengalese finch (*Lonchura striata var. domestica*) and examined whether neurons in Area X shows singing-related and auditory-related activity by recording multiple single-unit activity in freely behaving condition.

We found two types of audio-vocal mirror neurons in Area X that exhibited both singing and auditory-related activity. One type of neurons fired at the specific syllable timing when a bird was singing and when it was listening to BOS (Fig.1). Based on the firing property and spike waveform (Goldberg & Fee, 2010; Goldberg et al., 2010), these neurons were classified as putative striatum medium spiny neurons (MSNs). The other type of neurons was active when a bird was singing. The same neuron also showed increase in firing rate when a bird was listening to BOS during sleep, while the neuron did not show such auditory responses when a bird was awake. These neurons were classified as internal globus pallidus (GPi) neurons.

MSN is the input stage of Area X from HVC (Farries et al., 2005). Thus, MSN may exhibit similar audio-vocal mirror neuron properties by receiving inputs from HVCx neurons. On the other hand, GPi is the main output in Area X to the thalamic nucleus DLM (Farries et al., 2005). Thus, we assumed that GPi send to DLM both auditory and vocal-motor information necessary for vocal learning. However, GPi neurons did not show any auditory response while MSN showed BOS selective responses when a bird was awake. These results suggest that the neural transmission from MSN to GPi is gated off when a bird is awake.

In conclusion, we found audio-vocal mirror neurons in Area X and these are putative MSN and GPi neurons. We assume that the neurons integrate auditory and vocal-motor information in the cortico-basal ganglia circuitry necessary for vocal learning. We also suggested that the auditory information processing is modulated by arousal levels. Our findings support the idea that audio-vocal mirror neurons in cortico-basal ganglia circuitry are important in vocal learning that needs to integrate sensory and motor information. The mirror neurons that integrate auditory and vocal information may be a common neural substrate not only for vocal learning in birds but also for speech learning in humans, and this study provides new clues for considering the evolution of language.

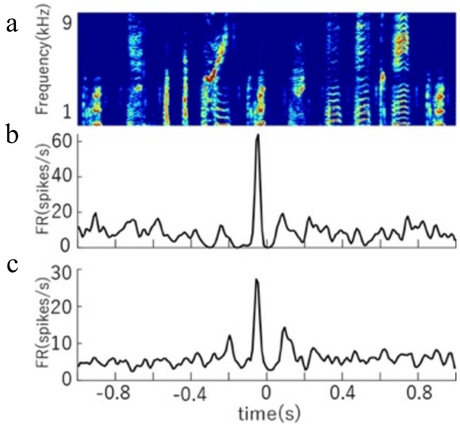


Figure 1. An example of firing pattern of audio-vocal mirror neuron in adult Bengalese finch basal ganglia nucleus Area X. (a) Sonagram. (b) Mean firing rate (FR) when a bird was singing. (c) Mean FR when a bird was listening during wakefulness. Note that the phasic firing occurs at the time of specific syllable during singing and listening to the song.

## Acknowledgements

This work was supported by MEXT/JSPS KAKENHI Grant Number #4903, JP17H06380, and JP20H00105.

## References

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, *22*, 567–631.

Farries, M. A., Ding, L., & Perkel, D. J. (2005). Evidence for "direct" and "indirect" pathways through the song system basal ganglia. *The Journal of Comparative Neurology*, *484*(1), 93–104.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in genie: a case of language acquisition beyond the "critical period." *Brain and Language*, *1*(1), 81–107.

Goldberg, J. H., Adler, A., Bergman, H., & Fee, M. S. (2010). Singing-Related Neural Activity Distinguishes Two Putative Pallidal Cell Types in the Songbird Basal Ganglia: Comparison to the Primate Internal and External Pallidal Segments. *Journal of Neuroscience*, *30*(20), 7088–7098.

Goldberg, J. H., & Fee, M. S. (2010). Singing-Related Neural Activity Distinguishes Four Classes of Putative Striatal Neurons in the Songbird Basal Ganglia. *Journal of Neurophysiology*, *103*(4), 2002–2014.

Hessler, N. A., & Okanoya, K. (2018). Physiological identification of cortico-striatal projection neurons for song control in Bengalese finches. *Behavioural Brain Research*, *349*, 37–41.

Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature*, *433*(7026), 638–643.

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology*, *71*(2, Pt.2), 1–25.

Prather, J. F., Okanoya, K., & Bolhuis, J. J. (2017). Brains for birds and babies: Neural parallels between birdsong and speech acquisition. *Neuroscience & Biobehavioral Reviews*, *81*, 225–237.

Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory–vocal mirroring in neurons for learned vocal communication. *Nature*, *451*(7176), 305–310.

# HOW MUCH LANGUAGE USE IS ACTUALLY ON SOCIAL TOPICS?

ANNA SZALA*[1,2], MAREK PLACIŃSKI[1], ALEKSANDRA PONIEWIERSKA[1], ALEKSANDRA SZCZEPAŃSKA[1] and SŁAWOMIR WACEWICZ[1]

*Corresponding Author: szala@oakland.edu
[1]Centre of Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland
[2]Department of Psychology, Oakland University, Rochester, MI, USA

## 1. Introduction

How much language use is on social topics? Dunbar and colleagues (1997) claimed that "socially relevant topics" accounted for about two-thirds of time spent on conversation, and later studies largely confirmed this pattern, including cross-culturally (e.g., Kubik et al., 2018). In addition to its considerable popular impact, Dunbar et al.'s study broke new theoretical grounds, helping motivate the evolutionary Social Brain theory and the influential "gossip" theory of language origins (however, the term "gossip" is problematic and should be used with caution – see esp., Dores Cruz et al., 2021).

Here, we revisit Dunbar's question about the proportion of conversational time devoted to social topics. In this effort, we are motivated by the substantial limitations of both the dataset and approach of the original study. Dunbar et al. (1997) relied on a relatively small number of conversations (N = 45), collected exclusively in open public environments, between a sample of participants with a very limited demographic and geographic distribution. Secondly, their definition and operationalisation of "social topics" lacked clarity and sufficient connection to the proposed adaptive functions of language. For example, Dunbar et al.'s classification of the categories of sport/leisure, culture/art/music, or politics as non-social appears arbitrary in the light of our own data (see below) and grouping "personal experiences" within the larger category of social topics does not appear to demonstrate a straightforward link to the proposed adaptive social functions of language that motivate the evolutionary conclusions drawn from this study.

## 2. Materials and methods

We used Spokes (Pęzik, 2014), a corpus of Polish informal, casual conversations (N = 668; over 2.6 million word tokens) with speakers from a large variety of Polish social backgrounds. Mindful of the limitations of Dunbar's original approach, we re-defined "social topics" as conversations concerning 1) people that at least one participant knew in person (e.g., neighbour), 2) people they did not know in person (e.g., musician), and 3) social groups they either knew or did not know in person (e.g., classmates, football teams). As a consequential decision, content related to the speakers' personal experiences which did not involve others was considered non-social. We randomly selected 56 conversations (ca. 10% of the dataset excluding the training set), which were independently assessed by two judges, who rated each text line in each transcript as concerning a social or non-social topic.

## 3. Results

The expert judges converged on 85% of their judgments (moderate interrater agreement; $\kappa = 0.68$). Portions of conversations classified differently by the judges were removed from the analysis. The total number of word tokens was 52436 for social topics (50.9%) and 50528 for non-social topics (49.1%). We fitted age and gender to a logistic regression model to check how they link with conversing on social topics. Age positively correlated with conversing on social topics ($\beta = 0.015$, $p < 0.001$); gender did not show an effect but a trend was for males to speak less on social topics ($\beta = -0.073$, $p = 0.09$).

## 4. Conclusions

We found a roughly equal share of conversation devoted to social versus non-social topics, which differs considerably from the roughly 2/3 reported by Dunbar et al. (1997). One key difference was our classification of "personal experiences" (content limited to the individuals involved in the conversation) as non-social. Unsurprisingly, the proportion of social topics in conversation is very sensitive to the definition applied, and since developing a standard multipurpose definition does not appear feasible (cf. e.g. phatic communication is technically not on "social topics"), further research should, in particular, compare the different definitions of "social topics" that are directly rooted in the different socially adaptive functions of language.

## Acknowledgements

This research was supported by the Polish National Science Centre under grant agreement UMO-2019/34/E/HS2/00248.

## References

Dores Cruz, T. D., Nieper, A. S., Testori, M., Martinescu, E., & Beersma, B. (2021). An Integrative Definition and Framework to Study Gossip. *Group & Organization Management, 46*(2), 252–285. DOI: 10.1177/1059601121992887

Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature, 8*, 231–246. DOI: 10.1007/BF02912493

Kubik, B. G., Pápay, B. T., Takács, K., & Galántai, J. (2018). The Gossip Hypothesis. [Prezi slides]. European Research Council. www.prezi.com/bs2dnr14_j2x/the-gossip-hypothesis.

Pęzik, P. (2014, October 24–25). Spokes – a search and exploration service for conversational corpus data. [Selected Papers]. CLARIN 2014: Soesterberg, The Netherlands. www.ep.liu.se/ecp/116/009/ecp15116009.pdf.

# WHAT LINGUISTICS TEXTBOOKS TEACH ABOUT LANGUAGE EVOLUTION

ALEKSANDRA SZCZEPAŃSKA*[1], ALEKSANDRA PONIEWIERSKA[1], MICHAEL PLEYER[1] , PRZEMYSŁAW ŻYWICZYŃSKI[1] and SŁAWOMIR WACEWICZ[1]

*Corresponding Author: szczepansk29@gmail.com
[1]Center for Language Evolution Studies, Nicolaus Copernicus University in Toruń, Toruń, Poland

## 1. Introduction

In 2016, Wacewicz et al. (2016) conducted a preliminary study of topics related to language evolution research as presented in 14 introductory-level linguistic textbooks, finding that in contrast to areas such as language acquisition, language change or language and the brain, language evolution received insufficient attention, often based on out-dated and inadequate conceptual frameworks. Here, we further investigate this issue with a first-of-its-kind, large computer-assisted qualitative study, in which we inspected 18 textbooks for all content related to the evolutionary emergence of language and its uniqueness in nature, in order to evaluate its thematic scope, selection of topics, theories covered, researchers cited, structural soundness, currency, and factual accuracy.

## 2. Materials and methods

Our dataset consisted of the most recent editions of 18 introductory textbooks to linguistics, selected based on the popularity measures *OS Appearances* and *OS Score* on the Open Syllabus Project (opensyllabus.org). We used the qualitative data analysis software nVivo 1.3 to mark and code any content related to the origins of language and its status among animal communication systems, appearing anywhere in the content of each book, including boxes, footnotes, captions, or exercises (but not indexes or references). We followed a 5-step coding procedure consisting of training (two language evolution experts instructed two expert coders), individual coding by the expert coders, consensus discussion of all coded passages by the coders, review by two language evolution experts, and consensus discussion by all four experts. This resulted in a hierarchy of 462 codes in the main thematic groups: Animal Communication (AC), Language Evolution (LE), and Researchers (R).

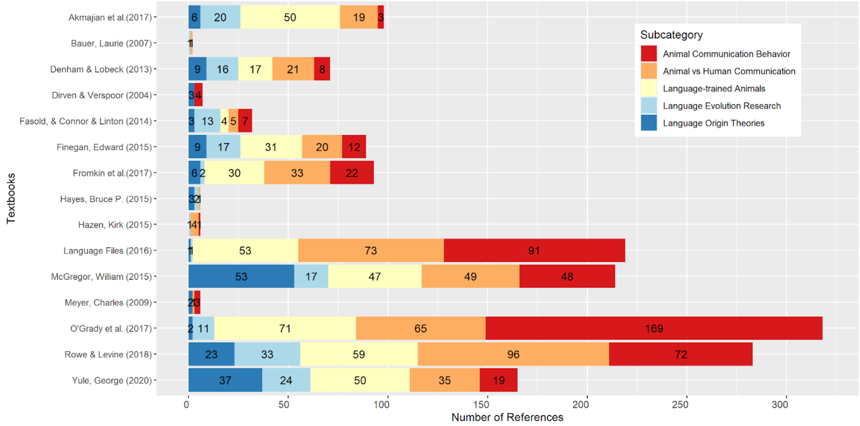


Fig. 1. Numbers of references to the main thematic categories of interest in 15 introductions to linguistics. The remaining three textbooks did not contain any references to either Animal Communication or Language Evolution.

## 3. Results and discussion

Overall, we found that the content of interest lacks a canonical representation across the textbooks (e.g. of the 189 researchers whose work was cited, only 17 were cited in 5 or more textbooks). The coverage of animal communication was relatively broad, with some recurring classic examples, such as vervet monkeys or honeybees. As its main problem, we see overreliance on the outdated framework of “design features” (11 textbooks, 183 references), to the exclusion of other important frameworks such as signaling theory (0 references) or phenomena such as turn-taking (3 books, 2 references). In contrast, the coverage of topics related to language origins and evolution was much less extensive and systematic, and suffered from an overrepresentation of content whose value is almost exclusively historical (i.e. creation myths, “bow-wow” theories), at the expense of current research in this fast-paced field; resulting in almost complete exclusion of topics such as cultural evolution (1 book, 3 references) or tool use (3 books, 7 references). We found the textbooks to be mostly factually correct, with a total of 60 problematic passages of different dimensions found in all textbooks (38 for AC and 22 for LE). We conclude with a detailed discussion of the results and several recommendations for future editions of the textbooks, most importantly changing the proportion of historical content to recent results, and including signalling theory and cultural evolution.

## Acknowledgements

This research was supported by the Polish National Science Centre under grant agreement UMO-2019/34/E/HS2/00248.

## References

Wacewicz S., Żywiczyński P. and Jasiński A. (2016). "Language Evolution And Language Origins In Teaching Linguistics At The University Level". In S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér & T. Verhoef (eds.) *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11).*

# INFERRING THE TEMPORAL AND SPATIAL EVOLUTION OF ACCENT SYSTEMS IN JAPANESE DIALECTS: A PHYLOGEOGRAPHIC APPROACH

Takuya Takahshi*[1], Ayaka Onohara[2], and Yasuo Ihara[3]

*Corresponding Author: takuya.14966@gmail.com
[1]Meiji Institute for Advanced Study of Mathematical Sciences, Tokyo, Japan
[2]Center for Statistics and Information, Rikkyo University, Tokyo, Japan
[3] Department of Biological Sciences, the University of Tokyo, Tokyo, Japan

## 1. Abstract

Accent systems of Japanese dialects over Japan's mainland are characterized by pitch changes, which are further subdivided into two phonological features; pitch accent and word tone (Hayata, 1999; see also Labrune, 2012; Uwano, 2012). As for the former, a majority of the dialects accentually distinguish words, that is, words are distinguished by the mora where the voice-pitch changes (McCawley, 1968; Uwano, 2012). For example, consider three words "hashi" (chopsticks), "hashi" (bridge), and "hashi" (edge) followed by the topic marker "ga." In Tokyo dialect, the phrase "hashiga" (mora structure: ha-shi-ga) is pronounced in a high-low-low manner for "chopsticks," low-high-low for "bridge," and low-high-high for "edge". These words are thus distinguished by the position of pitch accent, where the pitch drops, which is placed on the first and second morae for "chopsticks" and "bridge," respectively, but absent for "edge." On the other hand, "word tone" (Hayata, 1999) or "tonal register" (Uwano, 2012) concerns the change of voice pitch in a whole word, which characterizes the accent system of some regions such as the Kyoto and Kagoshima dialects, but is absent in other dialects. Utsugi (2007) describes that word tone (Hayata, 1999) is an almost equivalent concept to N-pattern accent (Uwano 2012).

Studying these two phonological features potentially gives insight into the ancestry of the Japanese language for the following reasons. First, the word tone has been pointed out to be a continuous feature of the Chinese tones (Hayata, 1999). Second, since pitch accent and word tone affect the whole phonological

system, they change only infrequently, and their low speed of replacement enables the trace of far past lineages and phylogenic reconstruction of regional dialects.

In the current research, we distinguish four accent patterns according to the presence/absence of pitch accent and word tone, aiming at elucidating how this variation was formed after the regional dialects had split from the common ancestor, in which both pitch accent and tone were present. In particular, the purpose of this research is two-fold: (1) inference of transition rate between accent patterns and (2) estimation of phylogenic tree of regional dialects based on their accent systems. To create empirical data, we randomly select 100 locations in Japan and specify the accent pattern at every location. To consider the spatial contact patterns between human groups, we develop a network of populations, which reflects the real geography and population density of Japan's mainland, and simulate the spread of accent patterns on the network (see Takahashi & Ihara, 2020). Simulating the genealogy of accent patterns, and applying a method borrowed from Bayesian phylogenetics (Felsenstein, 1981; Pagel & Meade, 2017), we obtain the posterior distribution of transition rates among accent patterns, the range of spatial interaction between populations, and the phylogenetic topology and divergence time of the present dialects.

Thus far, we have quantified the loss rates of pitch accent and word tone, but are unable to reconstruct a clear phylogenetic tree with a sufficiently high posterior clade probability. Future study needs to include more data.

## Acknowledgements

This research was supported by JSPS KAKENHI, grant numbers JP20J13493, JP18J00484, and JP17H06381. This research was also supported under the framework of the Japanese-Swiss Young Researchers' Exchange Programme of the Leading House Asia at ETH Zurich.

## References

Felsenstein J. (1981). Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. Journal of Molecular Evolution. 17: 368-376.
Hayata, T. (1999). Oncho no Taiporojii. Tokyo, Japan: Taishukan.
Labrune, Laurence. (2012). The phonology of Japanese. Oxford: Oxford University Press.
McCawley, J. (1968). The Phonological Component of a Grammar of Japanese. The Hague, Mouton.
Pagel M., Meade A. (2017). The deep history of the number words. Phil. Trans. R. Soc. B 373: 20160517.
Utsugi, A. (2007). The interplay between lexical and postlexical tonal phenomena and the prosodic structure in Masan / Changwon Korean, *ICPhS 2007 Satellite Meeting: Workshop on "Intonational Phonology: Understudied or Fieldwork Languages*," Saarbrücken, Germany.
Uwano, Z. (2012). Three types of accent kernels in Japanese, Lingua, 122(13).
Takahashi T., Ihara Y. (2020). Quantifying the spatial pattern of dialect words spreading from a central population. Journal of the Royal Society Interface 17, 20200335.

# CUMULATIVE CULTURAL EVOLUTION, POPULATION STRUCTURE AND THE ORIGIN OF COMBINATORIALITY IN HUMAN LANGUAGE

MONICA TAMARIZ*[1], SIMON KIRBY[2]

*Corresponding Author: m.tamariz@hw.ac.uk
[1]Department of Psychology, Heriot Watt University, Edinburgh, UK
[2]Centre for Language Evolution, University of Edinburgh, UK

## 1. Introduction

The emergence of combinatoriality (or duality of patterning) in human language is a central question for evolutionary linguistics. Combinatoriality is absent only in a handful of natural languages, such as ABSL, an emergent sign language that developed in a village setting (Sandler et al. 2011). The emergence of another design feature of language, compositionality, has been explained as a trade-off between learning biases for compressibility and communication pressure against ambiguity operating over cultural transmission (Kirby et al. 2015). We test two hypotheses: (H1) Combinatoriality also results from a similar trade-off. (H2) Population structure explains differences in combinatoriality across languages.

## 2. Methods

We constructed minimal languages that could nevertheless show combinatoriality. Each language had four atomic meanings, each mapped onto a 2-unit signal from the set $\{ac, bd, ad, bc, pr, ps, qr, qs\}$. The 4096 possible languages included 8 *Degenerate* languages, with the same signal for all meanings; 1632 *Holistic* languages with all distinct signals, and no reuse of units across signals; and 48 *Combinatorial* languages with all distinct signals, and maximum reuse of units across signals.

During learning, agents computed a Bayesian posterior probability distribution over the set of possible languages. The compressibility bias was modelled as a higher prior probability for languages with lower coding length. The disambiguation pressure was modelled as a bias against ambiguous languages during communication (Goodman & Frank 2016).

Two population dynamics were modelled. During a communicative interaction, with Vertical dynamics, akin to the situation in village sign language emergence, agents learn from the oldest agent in the population. With Horizontal dynamics, agents learn from other agents who are themselves learners. This is more like the situation in deaf-community sign languages, such as the case of emergence in schools for the deaf (Meir et al., 2010). The oldest agent was removed and a new one added every 20 interactions.

## 3. Results and discussion

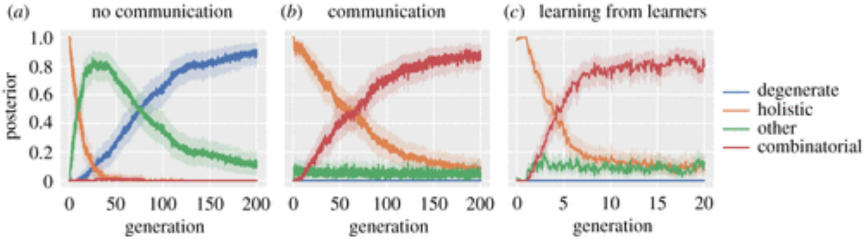


Figure 1. Posterior probability of each language type. In all cases agents have a bias for compressibility. (a) Vertical dynamics, no disambiguation pressure (200 generations). (b) Vertical dynamics, disambiguation pressure (200 generations). (c) Horizontal dynamics, disambiguation pressure (20 generations only).

With Vertical dynamics, under Compressibility bias alone (Fig. 1a), degenerate languages spread in the population; however, under Compressibility plus Disambiguation (Fig. 1b), combinatorial languages win out. This supports H1 that combinatoriality emerges as from a trade-off between compressibility and disambiguation pressures.

When learning from the oldest agent, combinatoriality evolves much slower than when agents learn from other learners. This supports H2 and suggests that population dynamics modulates the rate of evolution of combinatoriality, offering a new candidate explanation for why combinatoriality emerges more rapidly in deaf-community rather than in village sign languages.

## References

Goodman, N.D., Frank, M.C. (2016) Pragmatic language interpretation as probabilistic inference. *Trends Cogn. Sci. 20*, 818–829.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015) Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87–102.

Meir, I., Sandler, W., Padden, C. & Aronoff, M. 2010. Emerging sign languages. In M. Marschark, P. E. Spencer & P. E. Nathan (Eds.), *Oxford handbook of Deaf studies, language, and education* Vol. 2, 267–280. Oxford: Oxford University Press.

Sandler, W., Aronoff, M., Meir, I., Padden, C. (2011) The gradual emergence of phonological form in a new language. *Nat. Lang. Ling. Theory, 29*, 503–543.

# THE PREVALENCE OF SYSTEMATICITY IN INDO-EUROPEAN LANGUAGES

Nicole Tamer*[1,2], Paul Widmer[1,2]

*Corresponding Author: Nicole.tamer@uzh.ch
[1]Department of Comparative Language Science, University of Zurich, Zurich, Switzerland
[2]Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland

Human language has long been thought to be largely arbitrary, meaning that no intrinsic or logical connection exists between form and meaning (Hockett, 1960; de Saussure, 1983). Recent studies have revealed new insights by applying novel statistical methods, demonstrating the prevalence, and increasing relevance of non-arbitrary forms. These forms are commonly divided into iconicity, the resemblance-based mapping of form and meaning, (Winter & Perlman, 2021; Nielsen & Dingemanse, 2021) and systematicity, phonological cues predictive of grammatical categories (Raviv & Arnon, 2018; Pimentel et al., 2019; Nölle et al., 2018). An example of systematicity is the phonological distinction between open and closed word classes, the former commonly accept new words, whereas the latter does not, and within the open word class (Monaghan et al., 2007).

Because systematicity focuses on the relationships of large numbers of words to a small number of abstract categories, rather than the more frequent research on the relationships of single words to simple referential meanings, it has received less attention in the context of language evolution (Dingemanse et al., 2015). Languages tend to only exhibit subtle aspects of iconicity and these tend to decline over time in favour of more systematic or arbitrary forms (Little et al. 2017). Understanding the prevalence and mechanisms of systematicity is crucial to reveal more about its roots (cognitive advantages such as ease of processing, learnability, and acquisition advantages: Raviv et al., 2021; Monaghan et al., 2012; Fitneva et al., 2009) and role in the emergence and evolution of large grammatical and lexical inventories.

Corpus studies have revealed that some languages use subtle phonological and prosodic cues (e.g. stress, duration, voicing or phonotactics: Monaghan, et al., 2007) to differentiate between word classes and categories (Kelly 1992; Monaghan et al., 2005). However, previous studies included only a small sample heavily biased towards modern Western European languages. These factors constrain the ability to generalise these findings to other languages which limits broader inference about the role of systematicity in language evolution. To address this research gap and carry out an in-depth analysis of the prevalence of systematicity, we collected grammatical information of 30 modern and ancient languages extracted from language-specific grammars, corpora, and dictionaries. Specifically, we examined phonological cues towards grammatical categories of the first phoneme to capture the initial word recognition advantages provided by systematicity (Tamariz, 2008; Dingemanse et al. 2015).

A simple Shannon entropy was calculated within each grammatical category. The significance was tested by bootstrapping the original distribution of tokens to create new populations of entropy measures which provided information about expected entropy and the uncertainty of the measure.

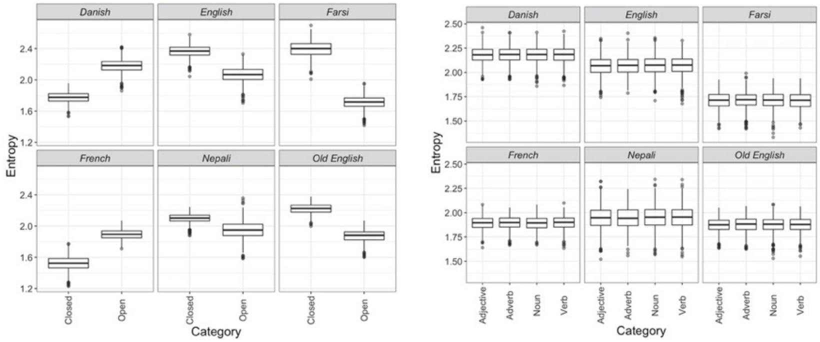


**Figure 1.** Entropy of the first word segment of closed and open word classes (left) with the highest differences and between grammatical categories of the open word class in a sample of languages.

We found significant differences in entropy measures of phoneme variability between open and closed word classes for every language across the Indo-European clades. No significant difference in variability was observed within the open word class itself when, for example, nouns were compared to verbs or adjectives with adverbs contrary to results of previous studies (Monaghan et al., 2007; Dingemanse et al., 2015).

We therefore conclude that a language-general pattern exists for differential variability in first word segment phoneme distribution between open and closed word classes. Intriguingly, this effect is not driven by variability between grammatical categories within the open word class contrary to previous findings. These findings demonstrate the critical and hitherto underestimated role that systematicity plays in the development and stability of certain grammatical macro-categories. Previous studies have not captured these aspects of systematicity, and these findings demonstrate how language behaves not only across the Indo-European language family but also within grammatical categories further highlighting the prevalence of non-arbitrariness in a sample of Indo-European languages.

## References

De Saussure, F. (1983). *Course in General Linguistics*, Open Court.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, *19*(10), 603-615.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of child language*, *36*(5), 967-997.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from?. *Cognitive science*, *31*(6), 961-987.

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, *203*(3), 88-97.

Kelly, P. (1992). Does the ear assist the eye in the long-term retention of lexis?. *IRAL: International Review of Applied Linguistics in Language Teaching*, *30*(2), 137.

Little, H., Rasilo, H., Van Der Ham, S., & Eryılmaz, K. (2017). Empirical approaches for investigating the origins of structure in speech. *Interaction Studies*, *18*(3), 330-351.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of experimental psychology: Learning, memory, and cognition*, *38*(5), 1152.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive psychology*, *55*(4), 259-305.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143-182.

Nielsen, A. K., & Dingemanse, M. (2021). Iconicity in word learning and beyond: A critical review. *Language and Speech*, *64*(1), 52-72.

Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, *181*, 93-104.

Pimentel, T., McCarthy, A. D., Blasi, D. E., Roark, B., & Cotterell, R. (2019). Meaning to form: Measuring systematicity as information. *arXiv preprint arXiv:1906.05906*.

Raviv, L., de Heer Kloots, M., & Meyer, A. (2021). What makes a language easy to learn? A preregistered study on how systematic structure and community size affect language learnability. *Cognition*, *210*, 104620.

Raviv, L., & Arnon, I. (2018). Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition*, *181*, 160-173.

Tamariz, M. (2008). Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *The Mental Lexicon*, *3*(2), 259-278.

Winter, B., & Perlman, M. (2021). Iconicity ratings really do measure iconicity, and they open a new window onto the nature of language. *Linguistics Vanguard*, *7*(1).

# INTEGRATIVE BRAIN ACTIVITY IN PROCESS OF VISUAL AND AUDIAL RECOGNITION OF IMITATIVE WORDS ON DIFFERENT DE-ICONIZATION STAGES

LIUBOV TKACHEVA*[1], MARIA FLAKSMAN*[2,3], YULIA LAVITSKAYA[4], YULIA SEDELKINA[4], ELIZAVETA KOROTAEVSKAYA[1], AND ANDREI NASLEDOV[1]

*Corresponding Authors: tkachewa.luba@gmail.com, Mariia.Flaksman@anglistik.uni-muenchen.de

[1]Department of Pedagogy and Pedagogical Psychology, Saint Petersburg State University, Saint Petersburg, Russia

[2]Department for English and American Studies, Ludwig Maximilian University, München, Germany

[3]Department of Foreign Languages, St. Petersburg Electrotechnical University, St. Petersburg, Russia

[4]Department of Foreign Languages and Linguo-Didactics, Saint Petersburg State University, Saint Petersburg

## 1. Iconicity and de-iconization as a process of language evolution

Iconicity is a relationship of resemblance (Peirce, 1940). In the lexicon it manifests itself in a form of imitative (onomatopoeic, sound-symbolic) words and ideophones (Hinton et al, 1994; Voelz et al, 2001). It is believed to be a design feature of the language (Voronin, 2005) that has played an important role in the proto-language (Moreno-Cabrera, 2012; Voronin, 2005). Indeed, in the absence of an established system of signs, it is likely that the first-ever elements of ‘language’ could have been based on the ‘natural’ form-meaning similarity. There is no possibility to study the proto-language *in vivo* to establish the role of lexical iconicity in the language origin and evolution. However, there is a possibility to conduct research on the imitative words existing in modern languages and to study *their* evolution. Flaksman (2017) established that imitative words undergo four de-iconization stages (SDs) and transform from ‘vivid’ iconic interjections on SD-1 (*ha-ha*!, *zzz*!) to words on SD-4 which have completely lost the original form-meaning correlation under the influence of phonetic and changes (*cliché*, once a sound of molten metal). Thus, de-iconization in a way can be considered as a model for evolution of a part of the language's lexicon. The study aims to study this process by means of EEG.

## 2. EEG markers of words recognition

There are several EEG markers connected with semantic processing, among them early event-related potential (ERP) components in the range 100–200 ms which are known to be sensitive to lexical frequency (Carreiras et al, 2005); N250, which is sensitive to orthographic similarity (Carreiras et al, 2009a) and the phonological status of the letters (Carreiras et al, 2009b); P300 component reflects the processes of distribution of arbitrary attention and stimulus categorization (Didoné et al., 2016); N400 component is associated with lexical–semantic access (Laszlo, & Armstrong, 2013).

## 3. Research aims, material, and methods

The *aim* of this research was to conduct an experimental EEG study on Russian words on different de-iconization stages and to establish whether there are differences in the brain activity while processing visual and audial stimuli of different nature.

### *3.1. Stimuli selection for the experiment*

The material for the research were 15 sound imitative (SI) words equally distributed into 3 groups according to the criterion of iconicity. The stimuli were preselected by means of a lexical decision task. To investigate the degrees of iconicity in word recognition the experiment by Sidhu et al. (2020) was partly replicated. The results revealed typical representatives of each group: explicit SI words – xlop (clap), čmok (smack), voj (howl), pisk (squeak), čix (sneeze); implicit SI words - žuk (bug, beetle), zud (itch), pux (fluff), xrjak (boar), gus′ (goose); and arbitrary, non-SI, words - vosk (wax), svod (vault), syp′ (rash), taz (basin, bowl), trost′ (cane).

### *3.2. Methods and procedure*

110 Russian adult participants took part in this experiment after signing up informed consent officially approved by ethical committee of S-Petersburg State University. The basic experiment was preceded by a preparatory step, which included the selection and validation of visual stimulus material for each lexical stimulus. Each participant was randomly presented with 60 words, 30 visually and 30 audibly. The task of the subjects was to identify the word and choose the

appropriate picture from the two proposed. EEG was recorded using Mitsar electroencephalograph (bandwidth, 0.5–70 Hz) with a 250Hz sampling rate for each channel. 19 monopolar leads were arranged symmetrically according to the International 10–20 System (Fp1, Fp2, Fp3, Fp4, F7, F8, C3, C4, Fz, Cz, Pz, T3, T4, T5, T6, P3, P4, O1, and O2), EOG was recorded. Visual and auditory ERP were calculated using Repeated Measures Analysis of Variance.

## 4. Results and discussion

It was found that there was no statistically significant difference between *visually* presented explicit-, implicit-SI words and non-SI words. However, statistically significant differences were obtained for *audibly* presented explicit SI words in contrast to implicit SI words at N400 ERP (p=0.014050) and implicit SI words in contrast to non-SI words at P300 endogenous evoked potential (p=0.043261). The first result corresponds to the fact that the N400 is larger for figurative language (Kutas, & Federmeier, 2011). The second one lets us speculate that explicit SI words demand more cognitive resources in process of audial recognition than non-SI words.

## 5. Conclusions

We assume that the results obtained indicate a specific brain response associated with directed attention in the process of cognitive decision-making task regarding explicit and implicit SI words presented audibly, which may reflect a higher level of cognitive complexity of identifying this type of stimuli. Explicit SI words, thus, are not only defined as extra-systemic and ‘archaic’ by means of linguistics, but also stand out according to our findings in EEG experiments.

## Acknowledgements

The research was supported by RFBR grant No. 20-013-00575 “Psychophysiological indicators of the perception of sound symbolic words in native and foreign languages”.

## References

Carreiras, M., Duñabeitia, J. A., & Molinaro, N. (2009a). Consonants and vowels contribute differently to visual word recognition: ERPs of relative position priming. *Cerebral Cortex,* 19(11): 2659–2670.

Carreiras, M., Gillon-Dowens, M., Vergara, M., & Perea, M. (2009b). Are vowels and consonants processed differently? Event-related potential evidence with a delayed letter paradigm. *Journal of cognitive neuroscience*, 21(2): 275–288.

Carreiras, M., Vergara, M., & Barber, H. (2005). Early event-related potential effects of syllabic processing during visual word recognition. *Journal of cognitive neuroscience*, 17(11): 1803–1817.

Didoné, D. D., et al. (2016). Auditory evoked potential P300 in adults: reference values. *Einstein (Sao Paulo, Brazil)*, 14(2): 208–212.

Flaksman, M. (2017a). Iconic treadmill hypothesis. the reasons behind continuous onomatopoeic coinage. In M. Bauer, A. Zirker, O. Fischer & Ch. Ljungberg (Eds.), *Dimensions of Iconicity* (pp. 15-38). Amsterdam: John Benjamins.

Hinton, L., Nichols, J., & Ohala J. J. (Eds.). (1994). *Sound symbolism*. Cambridge: Cambridge University Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62: 621–647.

Laszlo, S., & Armstrong, B.C. (2013). Applying the dynamics of postsynaptic potentials to individual units in simulation of temporally extended ERP reading data. In Proceedings of the 35th Annual Conference of the Cognitive Science Society (Knauff, M. et al., eds), 2826–2831, Cognitive Science Society.

Moreno-Cabrera, J. C. (2012). The role of sound symbolism in protolanguage: Some linguistic and archaeological speculations. *Theoria et Historia Scientiarum*, 9:115, 119-132.

Peirce, Ch. S. (1940). *The Philosophy of Peirce: Selected Writings* (ed. by J. Buchler). New York: Harcourt, Brace and Company.

Sidhu, D., Vigliocco, G., & Pexman, P. (2020). Effects of iconicity in lexical decision. *Language and cognition* 12(1): 164-181.

Voeltz, E. F. K. & Kilian-Hatz. Ch. (Eds.). (2001). *Ideophones* [Typological Studies in Language 44]. Amsterdam-Philadelphia: John Benjamins.

Voronin, S. V. (2005). *Iconicity. Glottogenesis. Semiosis*. St. Petersburg: St. Petersburg University Press.

# EMBEDDING PARALLELEPIPED IN CO-OCCURRENCE MATRIX: SIMULATION AND EMPIRICAL EVIDENCE

Takuma Torii*[1], Akihiro Maeda[1], and Shohei Hidaka*[1]

*Corresponding Authors: {tak.torii,shhidaka}@jaist.ac.jp
[1]Japan Advanced Institute of Science and Technology, Ishikawa, Japan

Recent natural language processing technologies are based on the vector space models of language, in which each word is represented with a vector in high dimensional space. One of the earliest successes using the vector space models is the four-term analogical reasoning task. A certain quadruple of word vectors forms a "parallelogram" in the vector space, with which the fourth word vector is inferred as an answer to a given triplet of query words. Despite the large body of successful applications of the vector space models, it has remained unknown what the parallelogram means. This study aims to reveal this mystery. Our analysis suggested that analogical reasoning is possible by decomposition of the bigram co-occurrence matrix, and we demonstrated the formation of a parallelepiped by creating a miniature corpus and its word vectors. This analysis and demonstration imply a sort of symmetry or exchangeability in word-word co-occurrence structure.

## 1. Introduction: Distributional Models of Language

'Evolution' of the language processing capability of the machine is dramatic in this decade. The accuracy of machine translation has reached the human level or perhaps more than educated non-native speakers. These successes of machine-learning language models have suggested how natural languages are organized.

Language is usually considered as an organized system that exhibits the capability of determining a class of words given the context of a word to be determined. This theoretical idea is called *the distributional hypothesis* (Harris, 1954). The distributional hypothesis postulates that words that occur in similar contexts tend to have similar meanings. For example, 'an apple' and 'a banana' both are allowed to appear in similar contexts, e.g., "she eats _____ every morning" and "_____ is a fruit". However, they may not appear in similar contexts of 'a bus' and 'a train', e.g., "she takes _____ home". When we think of the fill-in-the-blank problem "she eats _____ every morning", the words that refer to something edible, women like, and common in breakfast would be coming up with, like those specified by the context of the blank to be filled.

One of the pervasive methods to implement the distributional hypothesis is counting the co-occurrence of words in the pairs, triplets, or n-grams. Such naive co-occurrence counting has, however, a few technical issues: the combinatorial space of word pairs is too large to sample sufficiently (e.g., a bigram (pair) table has $10^{12}$ cells for $10^6$ word types), and it causes underestimation on the co-occurrence probability. Thus, one needs further *compressed* representations of the co-occurrence table, that compressed representations hopefully preserve the distributional structure of the words in the table and the language. Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997) is one of such earliest attempts. The underlying idea of LSA is that a sparse co-occurrence matrix $M$ can be approximated by vector representation of words, called *word vectors*. It has been demonstrated that word-vector algorithms can solve semantic tasks, although their performances were limited (see Lenci (2018) for review).

More recently, Mikolov et al. (2013) discovered that four-term analogy problems can be solved accurately by their artificial neural network called *skip-gram*, which is an instance of the *word2vec* class of models. Four-term analogy problem questions "what is d to c as b is to a?" denoted by, a : b :: c : d. Formally, the model needs to predict word d given the triple of query words a, b, and c. For example, the question, man : woman :: king : _____, should be answered with 'queen'. Importantly, the word2vec was not optimized to solve the four-term analogy questions, but it was optimized to predict the context words for each word. However, with the learned word vectors, e.g., $v_{\text{king}}, v_{\text{man}}, v_{\text{woman}}$, one can answer the analogy task by vector arithmetic $v_{\text{king}} - v_{\text{man}} + v_{\text{woman}} \approx v_{\text{queen}}$. Since analogical reasoning requires not only syntactic but also semantic aspects of language, their successes in the analogy task have been viewed as strong support for the distributional hypothesis. And since analogy was considered to be 'uniquely' humans, this discovery gave a strong impact on a variety of research fields.

To solve such analogical questions, word2vec needs to successfully extract latent and distributional structures of the language, which is represented in the vector form. Since then, researchers of related fields have been attracted to resolve this "mystery" of word2vec, e.g., (Levy et al., 2015; Arora et al., 2015; Hashimoto et al., 2016). Most of them have concluded that the emergence of parallelograms is due to the sophisticated learning algorithm. The current consensus (see, e.g., (Lenci, 2018)) is the conclusion by (Levy et al., 2015) that the analogy performance of word2vec can be explained as a result of a factorization of the PPMI (positive pointwise mutual information), one of the most popular preprocess of co-occurrence matrix in natural language processing (NLP).

In this paper, we take a different approach to the mystery of word2vec. We hypothesized that the word co-occurrence matrix itself, rather than some transformation of it such as PPMI, has sufficient information required to solve linguistic tasks. Namely, we take one of the simplest forms of implementation of the distributional hypothesis. This approach has been rarely taken in the existing literature.

Connecting the co-occurrence matrix to analogical parallelograms directly naturally leads constructive approach — simulation to test which type of co-occurrence may embed a parallelogram in the word vector space. Thus, we take the two types of approaches, data-driven analysis of co-occurrence matrix and constructive simulation creating and manipulating a small corpus.

In what follows, we briefly introduce the word2vec model in Section 2, followed by an analysis of a co-occurrence matrix in Section 3, and the constructive approach in Section 4. Lastly, we discuss future directions toward the understanding of the semantic nature of underlying word co-occurrence.

## 2. word2vec: The Word Embedding Algorithm

We briefly introduce the key ideas of word2vec, specifically of the skip-gram artificial neural network architecture. The skip-gram model consists of the three layers, $n$ input units, $d$ hidden units, and $n$ output units, where $n$ is the vocabulary size. Initially, every word $w$ in vocabulary $W$ is represented by a so-called one-hot vector $e_w$ of length $n$. Given a long sequence of words represented by one-hot vectors, the goal of optimization is to obtain a $d$ dimensional compressed representation $v_w$, called word vector, for every word $w \in W$, where $d \ll n$. Denote by $w_t$ a word at the position $t$ in the corpus. The skip-gram seeks the corpus to identify every subsequence $(w_{t-k}, \ldots, w_{t-1}, w_t, w_{t+1}, \ldots, w_{t+k})$, the $k$ preceding and $k$ following context words around the centre word $w_t$. The skip-gram model is trained, to optimize the latent word vectors $\{v_w\}_{w \in W}$, for each $w_t$ to predict their all context words $\vec{w}_t = (w_{t-k}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+k})$ simultaneously throughout the corpus. Mikolov et al. (2013) defined for the skip-gram model the conditional probability of occurring $y$ in the context of $x$ as follows:

$$P(y|x) = \frac{\exp(v_y \cdot v_x)}{\sum_{w \in W} \exp(v_w \cdot v_x)} \ , \tag{1}$$

where $v_y \cdot v_x$ is the inner product of word vectors $v_x$ and $v_y$.

Using a trained word2vec, Mikolov et al. (2013) demonstrated that it can solve their four-term analogy questions. Consider, for example, the problem, man : woman :: king : _____, and the correct answer is ‘queen’. Given the word vectors $v_{\text{man}}, v_{\text{woman}}, v_{\text{king}}$ for the cue words, decide the most likely word $y$ by calculating the cosine similarity measure $\text{cosine}(v_x, v_y) = \frac{v_x \cdot v_y}{\|v_x\| \, \|v_y\|}$ for all words $x$:

$$v_y = \underset{v_x \,:\, x \in W}{\text{argmax}} \ \text{cosine}(v_{\text{man}} - v_{\text{woman}} + v_{\text{king}}, \ v_x) \ . \tag{2}$$

It is defined correct, if the word vector $v_y$ is $v_{\text{queen}}$. The overall percentage of correct answers is about 66% for the 19,544 questions.

If any model answers correctly for a quadruple using Equation (2), these four word vectors need to form a parallelogram in the vector space. Indeed, Mikolov et al. (2013) graphically showed parallelograms in a lower-dimensional subspace.

## 3. Analogical Reasoning with Raw Co-occurrence Matrix

The past studies exploring the analogical reasoning based on the word2vec or others (Levy & Goldberg, 2014; Hashimoto et al., 2016; Arora et al., 2015) have essentially hypothesized and concluded that word2vec or other transformation such as PPMI is crucial to have a good analogy performance. In this study, however, we hypothesize that a raw co-occurrence matrix itself or its matrix decomposition would be sufficient for analogical reasoning.

### 3.1. *Method*

To test our hypothesis, we directly counted the frequencies of pairwise co-occurrence of all words in the English Wikipedia dump corpus 20171001. The text data contains approximately 7.9 billion words, of which 2.6 million words are unique. The window size for word pair counting was $k = 5$. Although we counted them all, algebraic operations using the full co-occurrence matrix were impossible due to our computational power. Hence, for the analogical task, we only used the sub-matrices composed of the top 1,000 (or 10,000) unique words, in addition to the 905 unique words in the question set. Denote this co-occurrence matrix by $M \in \mathbb{R}^{n \times n}_{\geq 0}$ with vocabulary size $n$.

In NLP, it is commonly recognized that application of singular-value decomposition (SVD) to the co-occurrence matrix improves performance of linguistic tasks. Technically, SVD is a method for decomposition of a real matrix $M$ of arbitrary finite size to the form $M = U \Sigma V^\top$, where matrix $U$ and $V$ are real orthogonal matrices and the diagonal matrix $\Sigma$ contains singular values in its diagonal elements. By taking the first $d$ dimensions, the $d$ dimensional word vectors for $n$ words are obtained as $U_d \, \Sigma_d^{1/2} \in \mathbb{R}^{n \times d}$. Since the word2vec was trained to construct 300 dimensional word vectors, $d = 300$ was used in this paper.

We trained our word2vec (skip-gram) model using the sample code of Python library Gensim (Rehurek & Sojka, 2010). We used instead our own preprocessed text data as described above. The window size $k = 5$ is the same. Only the words that occur more than or equal to 100 times in the corpus were used for training the model. The number of unique words was approximately 0.32 million.

### 3.2. *Results*

Figure 1 shows the performance for the four-term analogy task using the distributional models. As shown by Mikolov et al. (2013), the performance of word2vec is 66%. We treat this as a benchmark. For the models freq, the rows of the co-occurrence frequency matrix $M$ were directly used as word vectors. The models showed accuracy below 5%. For the models logfreq, the logarithms of the rows of $M$ were used as word vectors. By taking the logarithms, the model performances got significantly increased about 40% and 35%. We think the logarithm worked as a smoothing against the Zipf's law. This partially supports our hypothesis that information required to solve linguistic tasks is inside the corpus data. However, there is room for further improvement (could be) induced by word2vec. To eliminate this possibility, we applied SVD, a classic word embedding method,

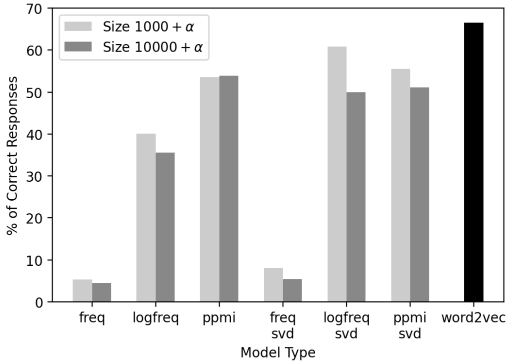


Figure 1. Four-term analogy performances of distributional models

to the log-frequency matrices of $M$. Since SVD is linear, although the word2vec is nonlinear, it would be helpful to resolve the mystery of word2vec. Surprisingly, the performances of the logfreq svd models are above 60% comparable to word2vec. This result supports the other half of our hypothesis that there is no latent structure that can be discovered only when using word2vec.

### 3.3. *Discussion: Why the Decomposed Co-occurrence Matrix Suffices*

If the original word2vec (skip-gram) were successfully trained, the word vectors $V \in \mathbb{R}^{n\times d}$ determines conditional co-occurrence probability matrix $P(y|x)$ in Equation (1). By taking logarithm, $V\,V^\top \in \mathbb{R}^{n\times n}$ is extracted (the normalizing term was ignored), and thus the skip-gram model could be viewed as an approximate matrix decomposition of the form $V\,V^\top \approx \tilde{M}$ for unknown $\tilde{M}$. Given the results in Section 3.2, it suggests that "up-to-$d^{\mathrm{th}}$-rank matrix decomposition of the logarithm of $M$" is essentially what the word2vec models do. This hypothesis differs from the previous study (Levy et al., 2015), which concluded that word2vec is equivalent to the PPMI-like smoothing, or a matrix decomposition of the PPMI-smoothed matrix of $M$. Our view, that word2vec as a co-occurrence matrix decomposition, can be viewed as one of the simplest and most straightforward implementations of the distributional hypothesis (Harris, 1954).

## 4. Constructive Approach to the Parallelograms

The analysis has suggested that there is a subspace of the co-occurrence matrix, in which a parallelogram is formed by a particular set of word vectors as each word may have multiple aspects. For example, *king* is more similar with *queen* on the *is-royal* axis, but is more similar with *man* on the *is-male* axis. Such a multi-aspect structure of the word *king* is supposed to be captured by a parallelepiped, rather

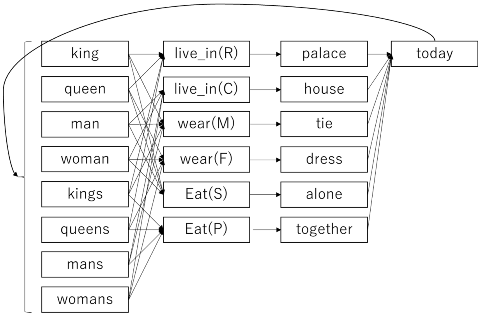


Figure 2. A hidden Markov model generating the 24 sentences in the toy corpus. Any hidden state X other than the verbs generates the word X by probability 1. For example, the state "king" generates the word "king". On the other hand, the two hidden states corresponding to the two verbs may generate the same word. For example, both states "live_in(R)" and "live_in(C)" generate the word "live_in".

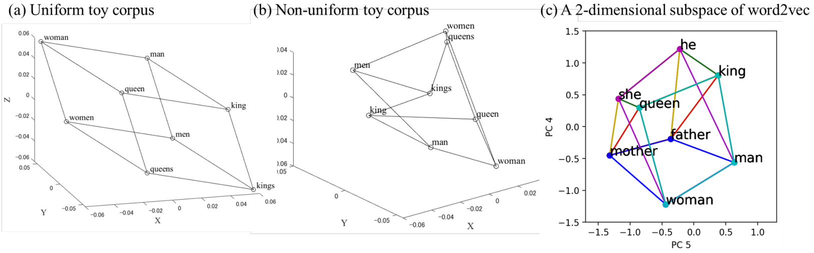
than a parallelogram. Although an analogy task tests a parallelogram, a collection of analogy tasks would test a parallelepiped or more complex geometric object.

In this section, we take a constructive approach to address how this parallelepiped structure is involved with the syntactic or semantic nature of a language. Specifically, we construct a small toy corpus, that forms an idealized parallelepiped structure among the word vectors, and analyzed what condition would be essential to form some parallelepiped of word vectors.

### 4.1. *Demonstrating Parallelepiped Embedded in Co-occurrence Matrix*

**Toy Corpus.** We created a corpus of 24 artificial sentences, which are not strictly grammatical, but with a minimal syntactic and semantic structure. Each of the sentences in this corpus consists of four words in the form of Subject-Verb-Object-Adverbial, such as "king live-in palace today". There are eight subjects, three verbs, six objects, and one adverb — in total 18 words. The corpus does not have all the possible sentences out of these 18 words, $144 = 8 \times 3 \times 6$, but it has only 24 sentences (Figure 2), which implicitly represents the hypothetical semantic relationship between underlying concepts to which these words refer.

First, we analyzed the co-occurrence matrix constructed for the toy corpus with each of the sentences generated by the equal probability $1/24$. In this case, the co-occurrence matrix (up to scale and permutation similarity) can be written with the two block matrices $C_0 \in \mathbb{R}^{8\times 10}$ and $C_1 \in \mathbb{R}^{10\times 10}$ by $C = \begin{pmatrix} \mathbf{0}_{8,8} & C_0 \\ C_0^\top & C_1 \end{pmatrix}$. Note that each row vector of the block matrix $C_0$ is the non-zero part of word vectors of the eight subject nouns. $C_0$ has the rank 4, and it lives in 3 dimensional affine space. Namely, there is some linearly independent basis of three vectors $b_0, b_1, b_2, b_3 \in \mathbb{R}^8$, such that $C_0 = (b_1, b_2, b_3)A + b_0\mathbf{1}_{1,10}$ with a unique matrix $A \in \mathbb{R}^{3\times 10}$ for each choice of the affine basis $B = (b_0, b_1, b_2, b_3)$. Let $\bar{B} =$



(a) Uniform toy corpus (b) Non-uniform toy corpus (c) A 2-dimensional subspace of word2vec

Figure 3. (Non-)parallelepipeds embedded in the co-occurrence matrix of (a) uniform toy corpus, (b) non-uniform toy corpus, and (c) a natural corpus.

$(b_1, b_2, b_3)$ has $b_1, b_2, b_3 \in \mathbb{R}^6$ of non-zero vectors of the column vectors of $\bar{C}_0 := C_0^\top - \frac{1}{10}\mathbf{1}_{1,10}C_0^\top$. Then the three dimensional coordinates of the 8 points are given by the column vectors of $(\bar{B}^\top \bar{B})^{-1}(\bar{B}^\top \bar{C}_0)$, in which a "parallelepiped" is embedded (Figure 3(a)). Thus, this uniform toy corpus gives a sufficient condition or the existence of a way to embed a parallelepiped in the co-occurrence matrix.

### 4.2. *Symmetry Breaker Against parallelepiped*

It is also important to demonstrate on which condition the parallelepiped embedded in a co-occurrence matrix is *broken*, as such a demonstration gives a necessary condition for the parallelepiped formation. To do so, we consider a variation of the toy corpus, called non-uniform toy corpus, in which a certain randomly assigned probability $p_i$ to sample the $i^{\text{th}}$ sentence to build the co-occurrence matrix. Figure 3(b) shows the same set of the eight word vectors visualized in the same way as Figure 3(a), for a set of non-uniform random probabilities $p_i$. These eight word vectors form neither a parallelepiped nor parallelograms. As the only difference between the uniform and non-uniform toy corpus is their sampling probability, this result suggests that a certain symmetric relationship in the probability distributions is needed to hold the parallelepiped.

### 4.3. *A Parallelepiped in Natural Co-occurrence*

The demonstration with the toy corpus above suggests that a certain class of word vectors would form a parallelepiped relationship, if the class of vectors two or more show independent syntactic-semantic statistical regularities on its word usage. We test this prediction by searching whether such a parallelepiped for a class of word vectors embedded in a natural co-occurrence matrix (logfreq svd, size 1000). Figure 3(c) shows an example that we found in the set of 8-tuple word vectors of the question words in the Family category (Mikolov et al., 2013) visualized by a two-dimensional subspace of the principal component analysis. This confirms our prediction.

## 5. Conclusion

This study attempted to give a theoretical account of what the parallelogram means in the vector space model. Our analysis of the co-occurrence matrix suggests a sort of co-occurrence matrix decomposition can give such a parallelogram useful for analogical reasoning. This empirical observation leads us to a constructive approach to building a toy corpus that may or may not embed a parallelepiped in the co-occurrence matrix. This numerical simulation suggests that the parallelepiped is tightly related to a certain class of the sentence probability distribution, perhaps less restricted than uniform but more restricted than arbitrary.

The biolinguistic enterprise of seeking cognitive precursors to human language depends on hypotheses or views on the structure of language. Our 'parallelotope hypothesis' may provide yet another characterization of the structure of language: word representations being structured are at least utilized for analogical reasoning among words. This hypothesis makes a strong connection between the mental representation of words of a language and relational reasoning on words. This hypothesis may motivate comparative psychology research on precursors to language in terms of the ability of relational reasoning.

## Acknowledgements

This work is supported by JSPS KAKENHI JP 20H04994, JST PRESTO JPMJPR20C9, Japan.

## References

Arora, S., Li, Y., Liang, Y., Ma, T., & Risteski, A. (2015). A latent variable model approach to PMI-based word embeddings. *Transactions of the Association for Computational Linguistics (TACL)*, *4*, 385-399.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2–3), 146–162.

Hashimoto, T. B., Alvarez-Melis, D., & Jaakkola, T. S. (2016). Word embeddings as metric recovery in semantic spaces. *TACL*, *4*, 273–286.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory. *Psychological Review*, *104*(2), 211–240.

Lenci, A. (2018). Distributional models of word meaning. *The Annual Review of Linguistics*, *4*, 151–171.

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In *Proceedings of NIPS 2014* (pp. 2177–2185).

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *TACL*, *3*, 211-225.

Mikolov, T. et al. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS 2013* (pp. 3111–3119).

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC Workshop* (pp. 45–50).

# GROUP SIZE AND LINGUISTIC STRUCTURE IN SILENT GESTURE

Ross Towns*[1], Tessa Verhoef[1]

*Corresponding Author: r.j.towns@liacs.leidenuniv.nl
[1]Creative Intelligence Lab, Leiden Institute for Advanced Computer Science (LIACS),
Leiden University, Netherlands

Evidence from natural languages suggests that properties of social environments are instrumental in shaping linguistic features, with languages of smaller or more isolated communities exhibiting more complexity (Lupyan & Dale, 2010, Wray & Grace, 2007) and less regularity (Lupyan & Dale, 2010, Meir, Israel, Sandler, Padden & Aronoff, 2012). In a recent artificial language study, Raviv, Meyer and Lev-Ari (2019) found that participants communicating within larger groups produced more systematic languages. They attributed this to the (initially) more variable input received by people in larger groups, which produced cultural selection pressures favoring systematicity as a means of reducing the cognitive strain of communicating with multiple partners.

We set out to extend this *input variability* hypothesis to be more readily applicable to accounts of sign languages in relation to social structure (e.g., Meir et al., 2012). Manual communication offers considerable opportunities for iconic signaling, which is associated with increased transparency of novel signals (Thompson et al., 2009) and may therefore counteract an effect of group size on the emergence of systematic structure. To investigate this possibility, we asked hearing non-signers to communicate using pantomime in groups of differing sizes. Below we report the procedure and results for the first two rounds, since conditions were similar across all group sizes for these rounds only.

We conducted the experiment online, using the Jitsi video-conferencing platform. Participants ($n$ = 26) were all masters students taking part at home as part of an introductory course on language evolution. Overall, we tested 3 groups of 5 people, 2 groups of 4, and 1 group of 3. Participants took turns to communicate a selection of concepts, presented as typed English words for their groupmates to interpret. The stimulus set was adapted from Motamedi et al. (2019), such that each item corresponded to a *thematic category* (e.g., *Religion*, *Music*, *Food*) and one *functional category* (*Person*, *Object*, *Location* or *Action*). Thus, the concept *Singer* corresponds to the categories *of Music* and *Person*, *Concert Hall* corresponds to *Music* and *Location*, and so on.

Following Motamedi et al. (2019), recordings of participants' gestures were manually coded for *shape* and *number of hands* (e.g., *2hTakePhoto* for gestures in which both hands are used to mime taking a photograph) and *use of functional markers* (gestures shared across items within a given functional category that may be interpreted as indicating that category, e.g., pointing to one's chest to denote a person). Coded trials were then analyzed using two measures of efficiency, namely *utterance length* and *frequency of repetitions*, and two measures of systematicity: *entropy* and *proportion of functionally marked utterances*, both of which indicate the recombination of elements across trials. Mixed effects linear regression models (with participant and round as random effects) found a significant negative effect of group size on entropy ($\beta = -0.32$, $SE = 0.14$, $t = -2.26$, $p = 0.032$*), but no effect on the proportion of functionally marked utterances. Marginally significant positive effects of group size were found for utterance length ($\beta = 0.43$, $SE = 0.23$, $t = 1.88$, $p = 0.073$) and frequency of repetitions ($\beta = 0.43$, $SE = 0.22$, $t = 2.00$, $p = 0.058$).

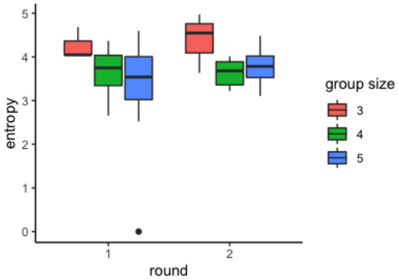


Figure 1. Mean entropy of participants' gestures. Error bars represent bootstrapped 95% confidence intervals

These results thus offer tentative support for the claim that larger groups produce more systematic gestures. However, the marginally significant results for gesture length and repetition frequency may suggest that larger group size hinders efficient communication. Though limited by our use of a small convenience sample, which was partly determined by pedagogical considerations, we argue that these findings provide a good basis for further investigation.

## References

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. Advances in Complex Systems, 15(3–4).

Meir, I., Israel, A., Sandler, W., Padden, C. A., & Aronoff, M. (2012). The influence of community on language structure. Linguistic Variation, 12(2), 247–291.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. Cognition, 192(April)

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. Proceedings of the Royal Society B: Biological Sciences, 286(1907), 20191262.

Thompson, R. L., Vinson, D. P., and Vigliocco, G. (2009). The link between form and meaning in American sign language: lexical processing effects. J. Exp. Psychol. Lang. Mem. Cogn. 35, 550–557.

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. Lingua, 117(3), 543–578

# THE CONDITION OF RECURSIVE COMBINATION IN THE EVOLUTION OF REINFORCEMENT LEARNING AGENTS

GENTA TOYA*[1], RIE ASANO[2], and TAKASHI HASHIMOTO[3]

*Corresponding Author: toyagent@protonmail.com
[1]Graduate School of Arts and Sciences, The University of Tokyo, Meguro, Japan
[2]Department of Systematic Musicology, University of Cologne, Cologne, Germany
[3]School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Japan

Language consists of multiple components with different evolutionary origins (Boeckx, 2013). Beside its central role in communication based on the capacity for intention sharing (Tomasello, 1999), another important component of language is its hierarchical structure. Human language requires not only a sequential operation but also a recursive operation generating hierarchical expressions by combining two elements, e.g., X and Y, into a unit {X, Y} which is again combined into another unit {Z, {X, Y}} (Chomsky, 1995). How and why did the recursive combination (hereafter RC) evolve? RC is significantly observed in object manipulation by humans (Greenfield, 1991) as well as by captive chimpanzees (Matsuzawa, 1991; Hayashi, 2007). Because object manipulation as a necessary skill for tool-use and -making can increase the fitness of an individual without cooperating with others, it is hypothesized that the RC of objects or actions evolved as a precursor of RC of lexical items or symbols (Fujita, 2017). We pursue this hypothesis of the evolutionary scenario that RC in motor control extended to RC in language in the course of human evolution.

In our simulation study, we model learning organisms as reinforcement learning agents (Sutton & Barto, 2018). By using agent-based modeling and evolutionary simulation, we investigate how and why RC evolved in learning organisms. In this simulation, agents, which are equipped with a reinforcement learning algorithm, explore and learn the process of tool making. Tool-making is implemented as a combination of elements through state transitions based on Q-values. In addition, the hyperparameters (learning rate $\alpha$, exploration rate $\varepsilon$, and time discount rate $\gamma$) of the neural network encoding the reinforcement learning and Q-values are explored by the genetic algorithm. The way agents make tools can be classified into two categories. One is the non-RC type production, in which elements are sequentially combined into a single object. The other is RC type production, where the combined object is re-combined with another object using a stack that can be acquired evolutionarily. The varieties of products that can be made with either production method remain the same. We set the reward function of the reinforcement learning algorithm so that the agents are rewarded more for making novel products. This corresponds to the phenomena that the invention of tools allows access to new resources (Arthur, 2009). We define the fitness function of the genetic algorithm so that the total rewards of an agent in a generation is discounted by the depth of the stack as a cost. In other words, this is a more favorable setting for non-RC than for RC because RC is more costly due to the additional use of the stack.

We found out that critical parameters for the emergence of RC include the cost of the stack, the reward discount rate, element types, and product length. If the cost of the stack, which is necessary for RC, is low, RC emerges because using RC in addition to the reinforcement learning is more advantageous than using reinforcement learning only. If the reward discount rate of producing the same product is low, it is less adaptive to produce as diverse products as possible. In this case, RC emerges because reinforcement learning with a high exploration rate ε is a less valuable strategy (Figure 1). If the types of elements are more and the product length is short, RC emerges because it is more effective than reinforcement learning with a low time discount rate γ.

In sum, in solving the exploration and exploitation problem, RC emerges in an environment where exploitation and exploration should be balanced. Moreover, the environment, in which attending to multiple states (i.e., to both workspace and stack) is not costly, facilitates the emergence of RC.

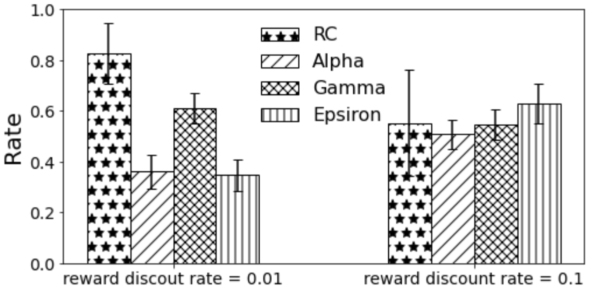


Figure 1. Reward discount rate dependency of the population share of RC users, hyperparameter of Q-learning (Alpha_QL, Gamma_QL, Epsiron_QL) at the 1,000th generation (average of 50 trials).

## References

Arthur, B. W. (2009). *The nature of technology: What it is and how it evolves.* The Free Press, NY: Simon & Schuster.

Boeckx, C. (2013). Merge: Biolinguistic considerations. *English Linguistics 30*, 463-484.

Chomsky, N. (1995). Bare Phrase Structure. In H. Campos. & P. Kempchinsky (Eds.), *Evolution and Revolution in Linguistic Theory* (pp. 1-15). Washington DC: Georgetown University Press.

Fujita, K. (2017). On the parallel evolution of syntax and lexicon: A Merge-only view. *The Journal of Neurolinguistics*, *43*, 178–192.

Greenfield, P. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Science*, *14*, 531–595.

Hayashi, M. (2007). A New notation system of object manipulation in the nesting-cup task for chimpanzees and humans, *Cortex*, *43*, 308–318.

Matsuzawa, T. (1991). Nesting cups and metatools in chimpanzees. *Behavioral and Brain Science*, *14*: 570–571

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

# CAN PHONOLOGICAL STRUCTURE EMERGE THROUGH ITERATED IMITATIONS OF ENVIRONMENTAL SOUNDS?

Julia Trzeciakowska*[1], Elizabeth Qing Zhang[1], and Sławomir Wacewicz[1]

*Corresponding Author: julietrz09@gmail.com
[1]Center For Language Evolution Studies, Faculty of Humanities, Nicolaus Copernicus University, Toruń, Poland

## 1. Introduction

Experimental studies on the development of communication have tended to focus on the role of the visual modality, and in particular the iconic potential of gestures and other forms of bodily communication, for the *de novo* emergence of communication systems in the laboratory (cf. Fay et al. 2014; Goldin-Meadow 2016). This line of enquiry has been relatively recently complemented with a new empirical approach to the potential for form-meaning iconicity in the vocal-auditory modality. These empirical studies reveal that vocal iconicity is inherently present in the lexicons of spoken languages (Johansson et al. 2021). Naïve listeners are able to assign new iconic vocalizations to their 'meanings' in forced-choice paradigms (Perlman et al. 2015), to a large extent even cross-culturally (Ćwiek et al. 2021). This has been taken to suggest that the iconic potential of the vocal modality might have played a role in establishing the first forms of language-like communication.

Further, existing research (Edmiston et al. 2018) with speakers of English suggests that linguistic properties such as phonological words can emerge spontaneously through repeated imitations of environmental sounds; however, the universality of these findings has not been tested beyond the native speakers of a single language. Our study extends this approach to participants speaking natively another three languages, from two language families, Indo-European (Polish and German) and Sino-Tibetan (Chinese). Also, four new onomatopoeic sounds were included.

## 2. Methods

Data was collected in a classic iterated learning paradigm, in chains of 8 participants per chain: first participant imitated a 'seed' environmental sound – such as glass, zipper or sneezing – and that participant's output was played back as input to be imitated by the next participant in a chain, etc. (cf. 'Chinese whispers'). We collected imitations from native speakers of Polish (8 chains), German (8 chains), and Chinese (8 chains). If imitations become gradually more language-like with each iteration, then the last (8th) generation imitations should sound more wordlike (thus, be easier to transcribe) than the first generation imitations. We asked native speakers of each language to convert the 1st and 8th generation imitations into spelling (native orthography). In the case of Chinese, participants used pinyin, i.e., the official romanised spelling; we expected more orthographic similarity (computed with the Ratcliff Obershelp algorithm) between the transcriptions of the 8th than 1st generation imitations.

## 3. Results

Paired samples t-tests revealed a significant difference between the mean similarity of transcriptions of sounds from the 1st and 8th generations in the expected direction only for the sounds provided by Polish speakers ($_{\mathrm{PL}}\, t_{(7)}=-3.201$, $p = 0.015 < 0.05$), and no significant differences for the sounds produced by the speakers of German ($_{\mathrm{DE}}\, t_{(7)} =-1.094$, $p =0.310 > 0.05$) or Chinese ($_{\mathrm{ZH}}\, t_{(7)=}$ 0.392., $p =0.707 > 0.05$). In a battery of qualitative post-studies, we established that the results heavily depended on the performance of the first participant in each chain (1st round imitator).

## 4. Discussion

We have been able to replicate the results of Edmiston et al. (2018) only for the sample of Polish speakers, but not German or Chinese speakers. Minimally, this suggests that care should be taken when generalizing results obtained from speakers of a single language, since these can reflect language or culture-specific phenomena rather than being informative of language-universal phenomena that are the basis for formulating language-origins scenarios.

## References

Ćwiek, A., Fuchs, S., Draxler, C. et al. Novel vocalizations are understood across cultures. Sci Rep 11, 10108 (2021). https://doi.org/10.1038/s41598-021-89445-4

Edmiston, P., Perlman, M., & Lupyan, G. (2018). Repeated imitation makes human vocalizations more word like. Proceedings of the Royal Society B, 285, 20172709.

Fay et al. (2014). Creating a communication system from scratch: gesture beats vocalization hands down. Frontiers in Psychology, 5(345). ). https://doi.org/10.3389/fpsyg.2014.00354.

Goldin-Meadow, S. (2016). What the hands can tell us about language emergence. Psychonomic Bulletin & Review, 24(1), 1–6.

Johansson, E., J. & Carr, J., S. Kirby (2021). Cultural evolution leads to vocal iconicity in an iterated learning task. Journal of Language Evolution, Volume 6, Issue 1, January 2021, 1–25, https://doi.org/10.1093/jole/lzab001

Kirby, S. & H. Cornish, K. Smith (2006) Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language, PNAS,

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. Royal Society Open Science, 2(8), 150152–16

Perniss P, Vigliocco G. (2014) The bridge of iconicity: from a world of experience to the experience of language. Phil. Trans. R. Soc. B 369, 20130300.

# ENVIRONMENTAL AND SOCIAL FACTORS IN THE EMERGENCE OF SPATIAL LANGUAGE

Kristian Tylén[*1,2], Zahra Alinam[3], and Cordula Vesper[1,2]

[*]Corresponding Author: kristian@cc.au.dk
[1]Deparment of Linguistics, Cognitive Science and Semiotics,
Aarhus University, Aarhus, Denmark
[2]The Interacting Minds Centre, Aarhus University, Aarhus, Denmark
[3]Architecture & Urbanism Faculty, Tabriz Islamic Art University, Tabriz, Iran

## Abstract

Languages across the world vary in how they construe simple spatial relations (Majid et al., 2004). It is, however, far from clear what the source of this variation is. One suggestion is that variation is driven first and foremost by cultural and historical factors such as language contact, transmission, and stochastic drift. As languages are passed down from generation to generation, cognitive biases for structure, compressibility and expressibility shape the emergence of linguistic structure and conventions (Christiansen & Chater, 2008; Kirby et al., 2015). In turn, when a speaker acquires a particular natural language, the structure of that language can influence how she will distribute attention and conceptualize her surrounding giving rise to *linguistic relativity* effects (Bowerman, 1996; Levinson, 2003; Roberson et al., 2005).

Another suggestion is that variation evolves contingent on ecological and environmental factors (Lupyan & Dale, 2016; Winters et al., 2015). The structure of our surrounding environment (affordances) can thus come to bias which conceptual solutions are experientially salient and thus influence which linguistic structures are stabilized (Nölle et al., 2020; Tylén et al., 2013). Here, we present a model that attempts to integrate the two elements (linguistic relativity and the linguistic niche hypothesis) in one dynamical system and provide experimental evidence to test its predictions. The model assumes dynamical circular causality

between environmental affordances and cognitive construal enhanced through cultural transmission (Alinam et al., 2021).

In a preregistered experiment, we used Virtual Reality to study the interactive processes shaping these affordances and their implications focusing on the case of urban environments. Seventy-two participants performed a navigation task first following, and later creating, instructions to find a target object in an urban environment. In each session, a participant would search a virtual urban space for a target, return and make written instructions for the next participant about where to go and look for the target. The same procedure repeated in a transmission chain of 6 generations, where the instructions of the previous participant was passed on to the next. In order to investigate the contingencies between environmental affordances, conceptualization and linguistic interaction, half of the participants did the task in an urban space characterized by windy streets and salient colorful building textures (henceforth the "Barcelona" condition), while the other half did the task in an urban space characterized by straight perpendicular streets and buildings with desaturated color/texture (the "Manhattan" condition). Number of houses, street crossings and their relative position was kept constant between the condition.

In support of our predictions, we find that route descriptions differed between conditions. Descriptions in the Barcelona condition were generally more reliant on local salient vertical landmarks while participants in the Manhattan condition relied more on horizontal and cardinal cues. This suggest that participants are susceptible to the visuo-spatial affordances on their local environment when preparing their instructions supporting the linguistic niche hypothesis. However, more interestingly, these tendencies are enhanced over generations in the transmission chains. Reading the verbal instructions of the previous participant seems to bias the attention of the current participant towards particular cues in the environment. Searching for the target, participants pay more attention to those dimensions of the space profiled in the instructions. And in turn, when later writing their own instructions for the next participant, they seem to not only reproduce description expressing the attentional profile of received instructions, but to even enhance such biases over generations. This is indicative of a linguistic relativity effect (at least according to a weak version of this).

Together, our results point to the mutually enhancing influences of (non-linguistic) environmental affordances and language in guiding attention and conceptualization, and can inform discussions of how linguistic structure emerge contingent on environmental structure.

## Acknowledgements

The authors would like to thank Astrid Endrup Iversen, Amalie Lysgaard Andersen, and Annamaya Halskov-Jensen for assistance with data collections. This study was supported by seed funding from the Interacting Mind Centre.

## References

Alinam, Z., Tylén, K., Pirbabaei, M. T., & Gharehbaglou, M. (2021). Cognitive-Cultural Looping Mechanism of Urban Space Conceptualization. *Integrative Psychological and Behavioral Science*, 1–19.

Bowerman, M. (1996). The origins of children's spatial semantic categories: Cognitive versus linguistic determinants. In J. J. Gumperz Levinson, S. C. (Ed.), *Rethinking Linguistic Relativity*. Cambridge University Press.

Christiansen, M., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(05), 1–70.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Levinson, S. C. (2003). *Space in Language and Cognition*. Cambridge University Press.

Lupyan, G., & Dale, R. (2016). Why Are There Different Languages? The Role of Adaptation in Linguistic Diversity. *Trends in Cognitive Sciences*, *20*(9), 649–660.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114.

Nölle, J., Fusaroli, R., Mills, G. J., & Tylén, K. (2020). Language as shaped by the environment: Linguistic construal in a collaborative spatial task. *Palgrave Communications*, *6*(1), 1–10.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411.

Tylén, K., Fusaroli, R., Bundgaard, P. F., & Østergaard, S. (2013). Making sense together: A dynamical account of linguistic meaning-making. *Semiotica*, *2013*(194), 39–62.

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, *7*(3), 415–449.

# INITIATING LEARNED VOCALIZATION IS PRECEDED BY LONG RAMPING IN NEURAL ACTIVITY IN BASAL GANGLIA IN A SPECIES OF SONGBIRD

SACHIO UMEMOTO*[1,2], SHIN YANAGIHARA[1,3], and KAZUO OKANOYA [1,3,4]

*Corresponding Author: umemoto@m.u-tokyo.ac.jp
[1] Graduate School of Arts and Sciences, The University of Tokyo, Japan

[2] Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo, Japan.

[3] Advanced Comprehensive Research Organization, Teikyo University, Japan
[4] RIKEN Center for Brain Science, RIKEN, Japan

Learning is essential to speak language. Humans are highly motivated to speak and improve it especially in their childhood. After critical period of language learning, they can still acquire new vocabulary, and even second language. It enables humans to sophisticate abstract and hierarchical language and contributes to communication among or between groups of large size. As well as the capacity of learning, motivation to learn language during their lifetime is needed on improvement of speaking language without being praised or rewarded, especially in their adulthood. It suggests that language learning should be motivated by internal system, and investigating mechanism of motivation of learning would help to understand how human language is maintained and evolved. In the perspective of motivation to maintain learned vocalization, vocal-learners are expected to have common system to vocalize frequently. As well as humans, male songbirds conduct vocal learning both as a child and an adult. Although their song is used to attract females, they also sing undirected song (US), and it has a role in maintain their syntax and acoustic features (Sakata & Brainard, 2006). It suggests that male songbirds are good model to investigate internal system to motivate vocalization. Anterior Forebrain Pathway (AFP), a cortico-basal ganglia-thalamo-cortical loop in songbird, is suggested to learn and maintain US (Brainard, 2004; Ölveczky et al., 2005; Hoffmann et al., 2016; Gadagkar et al., 2016). In AFP, neural activity in Area X, basal ganglia in songbird, is expected to reflect both preparatory activity and motivational signal because it receives premotor and dopaminergic input, and was reported on dopamine concentration during singing US (Sasaki et al., 2006). In order to investigate neural activity reflecting motivation in learned vocalization, present study recorded preparatory activity in basal ganglia of male songbirds during US, and compared duration of the activity change in US and compared it to that in vocalizing innate call.

Male Java sparrows (Lonchura oryzivora) (n = 4) were used for neural recording. Manually-driven electrode was implanted in their Area X, and neural activity was recorded while they vocalized US and innate calls in a sound-proof box. Neural activity was calculated by firing rate (FR) of sorted unit from recorded data. US bout was composed of motif (main body), and preceding vocalizations. FR in seven 1 second bins before the onsets and one bin after the onsets were used for statistical analysis. FR in 7-6 seconds before the onset was regarded as spontaneous firing rate (SFR), and song-related neurons were decided by Wilcoxon's sign rank test between FR in the bin after onset and SFR. Ramping duration of preparatory activity was decided by the number of continuous bins significantly higher in FR than SFR from the onset. As to innate calls, trill calls, specific long call in Java sparrow, was used in activity analysis. FR in three 1 second bins before the onsets and one bin after the onsets were used for statistical test. 24 neurons were recorded, 16 neurons showed US related activity and 6 showed trill call related activity. Ramping duration of preparatory activity was 2.25 ± 1.56 seconds in US (bout: 2.58 ± 1.55 seconds, motif: 1.25 ± 1.09 seconds), and 1.33 ± 0.47 seconds in trill call. The results shows that longer activity change in basal ganglia before learned vocalization than innate call in songbird. It suggests that initiation of learned vocalization may cause longer change in internal state, and motivate activity increase in various area, which may enables individuals complex and skilled vocalization.

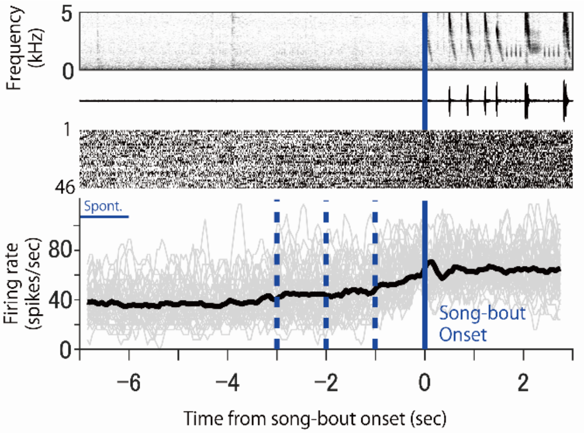


Figure: (Top) Spectrogram and waveform of US. (Middle) Raster plot of neural activity initiating US. (Bottom) Histogram of activity prior to US.

## References

Brainard, M. (2004). Contributions of the Anterior Forebrain Pathway to Vocal Plasticity. Annals of the New York Academy of Sciences, 1016(1), 377-394.

Gadagkar, V., Puzerey, P. A., Chen, R., Baird-Daniel, E., Farhang, A. R., & Goldberg, J. H. (2016). Dopamine Neurons Encode Performance Error in Singing Birds. *Science*, *354(6317)*, 1278-1282.

Hoffmann, L. A., Saravanan, V., Wood, A. N., He, L., & Sober, S. J. (2016). Dopaminergic contributions to vocal learning. *Journal of Neuroscience*, *36(7)*, 2176-2189.

Ölveczky, B. P., Andalman, A. S., & Fee, M. S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS biology*, *3(5)*, e153.

Sakata, J. T., & Brainard, M. S. (2006). Real-time contributions of auditory feedback to avian vocal motor control. *Journal of Neuroscience*, *26(38)*, 9619-9628.

Sasaki, A., Sotnikova, T. D., Gainetdinov, R. R., & Jarvis, E. D. (2006). Social context-dependent singing-regulated dopamine. *Journal of Neuroscience*, *26(35)*, 9010-9014.

# IMPLICIT COMMUNICATION OF WORD MEANING THROUGH CO-OCCURRENCE

Jeroen van Paridon*[1] and Gary Lupyan*[1]

*Corresponding Author: vanparidon@wisc.edu
[1]Department of Psychology, University of Wisconsin-Madison, USA

Communication requires people to align, at least in part, on what words mean. To accomplish this, language learners can observe what a word refers to. However, many words have referents that are abstract or otherwise hard to observe. In addition, many concrete words also have more abstract denotations. In the absence of direct referents, language learners can align word meanings to those of their language community by observing how words are used in context. This mechanism is underappreciated as a driver of semantic alignment across large language communities in which individual speakers are unlikely to ever interact directly. In three experiments, we demonstrate alignment between blind and sighted speakers for the semantic associations of color terms – whose direct referents can only be observed by the sighted participants – and demonstrate how a word embedding model can achieve this alignment by learning from word co-occurrence patterns.

Experiments in language evolution often focus on the transmission of structure rather than the transmission of semantics. For example, Kirby, Cornish, and Smith (2008) write that “each utterance has a dual purpose, carrying semantic content but also conveying information about its own construction. Upon hearing a sentence, a language learner uses the structure of that sentence to make new inferences about the language that produced it”. This is a foundational claim in the field of language evolution, and it has been repeatedly demonstrated empirically through, e.g., iterated learning experiments. In addition to conveying pragmatic meaning and information about its own construction, however, language also carries implicit information about its own lexical semantics in the form of word co-occurrences. For example, we might learn that “odd”, “strange”, and “weird” are related because because they are used in similar contexts. This is an underappreciated mechanism for language evolution, as it serves to align lexical semantics across speakers in a language community, which is vital for developing and maintaining a mutually intelligible lexicon.

Of course learning what words mean involves more than than tracking co-occurrences. Often, there is a direct referent present that the learner can observe. In some cases however, the language itself is the only source of information about lexical semantics that a language learner has access to. Blind people, for example, can only learn about the meanings of color words through language. We would expect, therefore, that they acquire those aspects of color word semantics that are implicitly conveyed in spoken and written language.

Recently, Saysani, Corballis, and Corballis (2021) showed that blind people's judgments resemble those of sighted people when asked to place color words along various dimensions, for example indicating where "red" and "green" fall on cold-hot, unripe-ripe and fast-slow continua. Given that blind people cannot directly observe that hot objects sometimes glow red or that unripe fruits and vegetables tend to be green, it perhaps seems obvious that any color associations they do have, they must learn from language (cf. Kim, Aheimer, Manrara, & Bedny, 2021). However, *how* color semantics are represented in spoken and written language – and to what extent language, rather than perception, can align semantic representations of colors between individuals – is not obvious. Are color semantics conveyed explicitly, e.g. through generic statements such as "green fruits are unripe"? Are they conveyed through simple co-occurrences, when a color word occurs adjacent to another word, e.g. "red hot coals"? Or are color semantics encoded in more complex semantic structures – a web of associations from which we can derive semantics of color terms?

## Experiment 1: Reanalysis of Saysani et al. (2021) data

### *Method*

#### *Participants*

Saysani et al. recruited 32 native speakers of New Zealand English, 20 of whom had normal, trichromatic vision and 12 of whom were congenitally blind with no residual vision. We recruited 130 additional sighted participants from the student participant pool at a large public university, speakers of American English.

#### *Design and procedure*

Participants were asked to rate each of nine color terms (red, orange, yellow, green, blue, brown, purple, black, and white) on 17 semantic dimensions, each defined by two antonyms placed at the poles of a seven-point Likert scale (happy–sad, calm–angry, submissive–aggressive, relaxed–tense, exciting–dull, selfless–jealous, active–passive, like–dislike, alive–dead, fast–slow, new–old, unripe–ripe, soft–hard, light–heavy, fresh–stale, clean–dirty, and cold–hot).

### *Results*

The main finding reported by Saysani et al. was that multidimensional scaling solutions were more variable between blind participants than between sighted participants. When we compared intraclass correlations (ICC) for the blind (.35, 95% CI [.29, .42]) and the sighted (.49, 95% CI [.43, .55]) groups, blind participants were indeed more variable than the sighted participants. At the same time, the responses of sighted and blind participants were remarkably similar (see Figure 1).

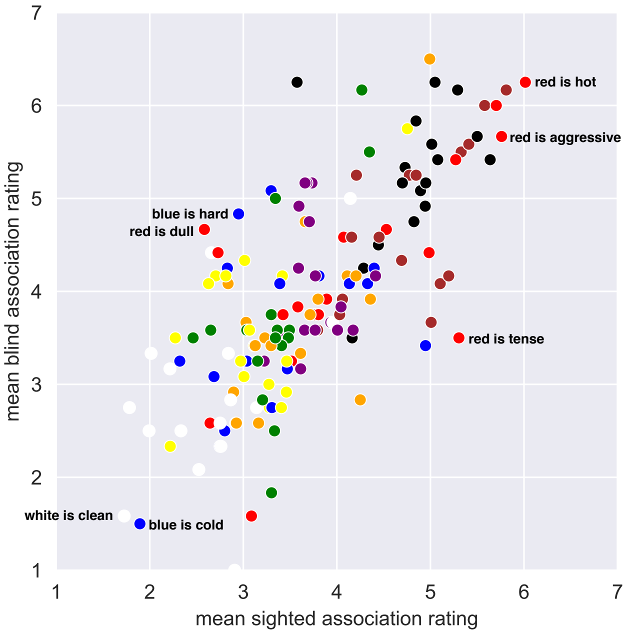


Figure 1. Blind and sighted participants' color-adjective association ratings from Experiment 1. Points on the diagonal from bottom-left to top-right represent perfect agreement between blind and sighted participants.

To understand how this aspect of color semantics may be represented in language, we relied on a fastText word embedding model (Bojanowski, Grave, Joulin, & Mikolov, 2017) trained on the fiction subcorpus of the Corpus of Contemporary American English (COCA-fiction). We projected the vector-representation of each color word onto a semantic dimension formed by the antonym pairs, e.g., *hot* and *cold*, and computing the cosine similarity between the color word vector and the axis vector. The projection for e.g. the color blue on the dimension cold-hot is then given by $\cos(\overrightarrow{\text{hot}} - \overrightarrow{\text{cold}}, \overrightarrow{\text{blue}})$ (see Grand, Blank, Pereira, & Fedorenko, 2018, for a discussion of this projection method). This provides us with a relative measure of word similarity, taken along the semantic dimension's axis, that we can use to predict human ratings of color associations.

Using a Bayesian linear mixed-effects model with weakly regularizing priors (Capretto et al., 2020), we regressed word embedding projections onto participants' color-adjective association ratings while adjusting for frequency and concreteness of the words forming each dimension. Color-adjective ratings (e.g., placing yellow closer to ripe than unripe) were predicted by word embedding projections, with a standardized effect size of .40 (95% CI [.37, .43]) for sighted participants and .33 (95% CI [.24, .41]) for blind participants.

### *Discussion*

Language is produced by people, most of whom have direct experiences of color. What is remarkable however, is that color information then becomes embedded in the statistics of language, enabling – in principle – someone who has no direct experience of color whatsoever to build up meaningful color semantics that can produce judgments quite similar to that of sighted people.

## Experiment 2: Where in language are color associations coming from?

So where do the embeddings "learn" their color semantics? One way of finding out is to remove the critical signal from the training corpus so that the resulting word embeddings no longer predict human judgments. In this experiment, we examined four potential sources of color-adjective associations:

(a) *First-order* co-occurrences: The occurrence of a color word and a semantic dimension word in the same sentence (e.g. "the fire was *red hot*"; color associations in these sentences can be explicit, but often are not).

(b) *Second-order* co-occurrences: The occurrence of color words and semantic dimension words in similar contexts (i.e. color words and semantic dimension words may not co-occur, but share words that they co-occur with, e.g. "Southern cooking uses *green* tomatoes" and "Southern cooking uses *unripe* tomatoes"). These sentences encompass nearly the entire corpus because some words (e.g. many function words) co-occur with every other word, which made removing all of them from the training corpus infeasible. More importantly, it rules out a strong form of the second-order co-occurrence hypothesis (i.e. *all* second-order co-occurrence relationships are informative), but it does not preclude a weaker form, where *some* second-order co-occurrences (e.g. the psychologically salient words from hypothesis (d)) are central to learning color-adjective associations.

(c) Co-occurrences between color words and words in the same semantic neighborhood as semantic dimension words: For example in "The forest was *white* with *snow*", *snow* is in the same semantic neighborhood as *cold*, which might lead to an association between *white* and *cold*). We identified semantic neighborhood words using cosine similarity between word embeddings and removed sentences containing any of the ten nearest neighbors of each color and dimension word from the corpus.

(d) Mediation by psychologically salient words: It is possible that color-adjective associations are mediated by specific words. For example, when placing *yellow* on the *unripe*-to-*ripe* dimension, people may think of a yellow and ripe banana. We do not know *a priori* which words mediate color-adjective associations, but we presented participants with color-adjective pairs (e.g., yellow-ripe, white-cold) and asked them to provide a word they associate with the pair. We then take the most common word for each pair and remove sentences containing those words from the training corpus.

Note that these sources of semantic information need not be mutually exclusive; words captured by (c) and (d) may overlap, and all of these words may be a subset of the words described by (b).

### ***Method***

#### *Participants*

We recruited 100 sighted participants from the student participant pool at a large public university who did not participate in previous color-adjective rating studies. Participants were presented with the color-adjective pairs and asked to generate a word that they associate with both. These associates were taken to be psychologically salient mediator words from hypothesis (d).

#### *Design and procedure*

To test each potential source of color-adjective associations, we removed it from the training corpus and then tested the predictive efficacy of embedding projections trained on the filtered corpus by using them to model the association ratings from Experiment 1.

### ***Results***

Removing first-order co-occurrences did not meaningfully reduce the effect size of the word embedding predictions. Removing nearest neighbors and especially removing participant-generated labels for color-adjective associations had a measurable impact however (see Figure 2 for estimated effect sizes).

### ***Discussion***

It is tempting to think that knowledge that blue is cold may come from sentences such as ”His lips were blue with cold”. However, removing such first-order co-occurrences had no measurable effect on the model’s ability to pull out human-like associations. In contrast, removing sentences containing psychologically salient mediators (e.g., ”ice” for cold-blue) reduced the signal substantially. This is especially surprising because the number of labels generated by at least two participants (the threshold for inclusion in our corpus filtering procedure) was only 242; on average less than one label per color-adjective pair.

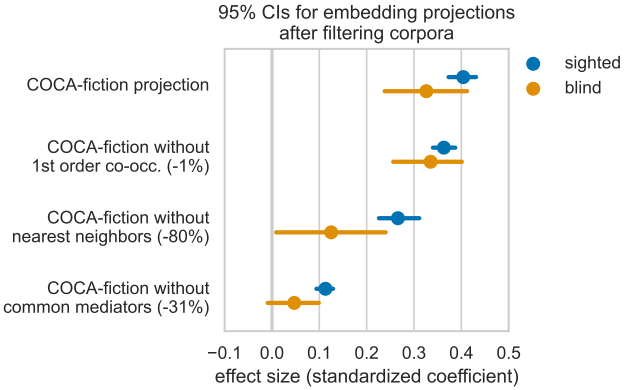


Figure 2. Estimated effects of word embedding projections in predicting blind and sighted participants' color-adjective association ratings. Percentage of training corpus removed by each manipulation is listed in parentheses.

## Experiment 3

To better understand what kinds of sentences were contributing to learning human-like color-adjective associations, we modified an embedding model to record color-adjective embedding projections at every single training step. This allows us to rank training examples in order of the impact they have on specific color-adjective projections (e.g. "blue" on the axis "hot"-"cold").

### *Method*

To measure the impact each individual sentence in the training corpus had on the embedding projections, we modified the word2vec (Mikolov, Chen, Corrado, & Dean, 2013) implementation included in the gensim Python package (Řehůřek & Sojka, 2010). The modified word2vec implementation computes and logs the embedding projections of interest after every training cycle (i.e. reading a training sentence, computing and back-propagating the error, and computing the updated embeddings). We then used the final embedding projections (after training is completed) as a reference and calculated how much each training sentence reduced the relative distance (between the previous projection and the final projection).

### *Results*

The sentences that most informed the final embeddings projections were (1) likely to contain either a dimension word (e.g., cold) or a color word, and (2) were likely to contain a color-adjective mediator produced in Experiment 2. For example, a highly informative sentence for moving "blue" toward "cold" is"The cold seaside air here has both a fishy and a piney sniff to it”. We can count up the occurrences of color and dimension words in the top 1000 most informative sentences for the “blue” and “hot”-“cold” pairing. We find 447 occurrences of “cold”, 326 occurrences of “hot”, and 303 occurrences of “blue”. Every sentence in the top 1000 contained at least one of these words, and only a few contained more than one. This suggests that the associations that underpin the projections are learned from specific second-order co-occurrences.

The most informative of these second-order co-occurrences are disproportionately mediated by words that participants in Experiment 2 named as salient labels for specific color associations (e.g. the association between ”yellow” and ”ripe” is mediated by salient label ”banana”). The top 1% of informative training sentences contains 2%–6% of the participant-provided mediator words in the training corpus, when aggregated by color.

### *Discussion*

Our results are strongly consistent with the model learning the color-adjective associations that inform the projections from second-order co-occurrence relationships. The higher prevalence of participant-proved mediator words for each color-adjective pair in the most informative training sentences demonstrates that participants were able to articulate with some degree of success which indirect (second-order) co-occurrence relationships are informative for the relationship between each given color and adjective (e.g. “white” and “cold”, mediated by the word “snow”).

## General discussion

In a language community where word meanings are always changing and where speakers cannot observe many words’ referents directly, how does a language learner align their understanding of word meanings to those of other speakers and the community at large?

One example of word meanings that have to be aligned without observing direct referents is blind people’s knowledge of color words. Blind people cannot directly perceive colors in their visual contexts, yet we found that their understanding of color associations is broadly aligned with that of sighted people, and that the color associations of both groups of participants could be predicted from word embedding projections. That these color associations can be learned from a corpus of written text by a model that learns from distributional information demonstrates how media, both spoken and written, could serve to align lexical semantics across a large language community in which most members never interact directly with each other. Communicating word meaning implicitly through co-occurrence also allows a language community to incrementally develop the meanings of abstract words–for which speakers cannot make use of referents–by scaffolding them on top of more concrete words.

Here, we used an adapted word embedding model to demonstrate exactly how co-occurrence information can be used by an associative learner to learn aspects of word meaning. We show that the core signal lies in second-order linkages mediated by a third word, e.g., the link between “ripe” and “red” being mediated by “tomato”. Large-scale semantic alignment and the mechanisms underpinning it are an under-explored topic in language evolution, but we believe that any comprehensive theory of language change needs to account for how language communities can maintain mutual intelligibility in the face of changing word meanings and varied access to direct perceptual information.

## Acknowledgements

This research was supported by NSF BCS grant 2020969, awarded to Gary Lupyan. We would like to thank Armin Saysani and Michael and Paul Corballis for making available the raw data used in Experiment 1.

## References

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *arXiv [Preprint]*. (`https://arxiv.org/abs/1607.04606`)

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2020). Bambi: A simple interface for fitting Bayesian linear models in Python. *arXiv [Preprint]*. (`https://arxiv.org/abs/2012.10754`)

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2018). Semantic projection: Recovering human knowledge of multiple, distinct object features from word embeddings. *arXiv [Preprint]*. (`https://arxiv.org/abs/1802.01241`)

Kim, J. S., Aheimer, B., Manrara, V. M., & Bedny, M. (2021). Shared understanding of color among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *118*(33).

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv [Preprint]*. (`https://arxiv.org/abs/1301.3781`)

Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop* (pp. 45–50). (`http://is.muni.cz/publication/884893/en`)

Saysani, A., Corballis, M. C., & Corballis, P. M. (2021). Seeing colour through language: Colour knowledge in the blind and sighted. *Visual Cognition*, 1–9.

# ON SCALING AND MODELS OF LANGUAGE EVOLUTION

REMI VAN TRIJP*[1]

*Corresponding Author: remi.vantrijp@sony.com
[1]Sony Computer Science Laboratories Paris, Paris, France

Computational models of language evolution offer important insights for explaining the emergence and evolution of human languages. However, such models have recently been criticized for being computationally intractable. The goal of this paper is to show that this criticism is misleading because it reduces all models of language evolution to only a specific subset of models that assume that the basic unit of cultural transmission is the language itself, which leads to astronomically large hypothesis spaces. In fact, there is already decades worth of computational modelling using the Language Game paradigm that has successfully addressed the issue of scaling by treating language as a complex adaptive system that spontaneously evolves as the side-effect of local communicative interactions. This paper explains why the Language Game method scales so well, and how it incorporates insights from constructivist usage-based learning and Relevance theory. It will illustrate the method through a Naming Game, supported by open-source code that readers can download, test and reuse for their own work.

## 1. Introduction

Computational models of language evolution have played an important role in exploring the origins and evolution of human languages ever since the late 1980s and early 1990s (e.g. Hurford, 1989; Steels, 1995). However, a recent complexity analysis by Woensdregt et al. (2021) suggests that models of language evolution, at least in their current formulation, are computationally intractable so they cannot be scaled up to more ecological scales involving tens of thousands of words.

While complexity analysis can offer useful insights for scaling a model, the criticism of Woensdregt et al. (2021) is misleading because it reduces all models of language evolution to one specific kind of iterated model based on Bayesian inference (e.g. Griffiths & Kalish, 2007). Other kinds of models are not only ignored, but also simply discarded based on the following two arguments:

> "[I]t is not clear that other models [...] would not run into the same wall of intractability. Moreover, the Bayesian formalism has the virtue of being able to model agents' epistemic states and transitions while remaining agnostic about the precise implementing mechanisms" (Woensdregt et al., 2021 p. 6).

The goal of this paper is to refute this conclusion by showing that there already exists decades worth of research using the Language Game methodology

(Steels, 1995, 2000) which has successfully addressed issues of scale by treating language as a complex adaptive system, and by drawing inspiration from constructivist usage-based language learning (Bybee, 2006; Goldberg, 2011) and Relevance Theory (Sperber & Wilson, 1986). This paper will explain why the Language Game method scales so well through a Naming Game (Steels, 1995; Baronchelli, Felici, Loreto, Caglioti, & Steels, 2006). Readers who are interested in running the Naming Game on their own computers can download the paper's supporting code for free as open-source software,[1] as well as the open-source framework Babel2 (Loetzsch, Wellens, De Beule, Bleys, & van Trijp, 2008) that has been used for the implementation, which can be downloaded at `https://gitlab.ai.vub.ac.be/ehai/babel-core`.

## 2. Illustrating the Problem of Scalability

Before turning to language games, it is important to understand the argument of computational intractability. Woensdregt et al. (2021) assume a Bayesian iterated learning model of cultural transmission in the style of Griffiths and Kalish (2007), in which an adult language user produces a number of utterances that are observed by a child learner, who forms hypotheses about which language could produce such utterances. Each hypothesis $h$ is a language, so the *hypothesis space* consists of all of the possible languages. At the end of a cycle, there is a generational turnover in which the child becomes the adult and a new learner is introduced.

Woensdregt et al. (2021) provide a complexity analysis that shows that such models are computationally intractable. They illustrate the idea with the following example: suppose that a language is a set of one-to-one mappings between *signals* and *referents*, then the hypothesis space of all possible languages consists of all possible signal-referent mappings, which amounts to $2^{\#signals \times \#\mathrm{re}ferents}$. Woensdregt et al. (2021) write that even for a toy language in which 50 signals exist for 25 referents, "learners need to consider all $2^{50 \times 25}$, about $1.9 \times 10^{376}$, possible languages" (p. 1). Increasing the number of signals and referents thus leads to a combinatorial explosion in the hypothesis space, which makes scaling impossible. In other words, such a model cannot be salvaged by faster computers or better implementations, because they face "a deeper theoretical issue" (p. 1).

Woensdregt et al. (2021) do not identify what exactly that deeper theoretical is, but the culprit seems obvious: the learner needs to consider an astronomically large hypothesis space, which is due to the fact that they have to consider the probabilities of all possible languages. This is a side-effect of the model's implicit assumption that the basic unit of cultural transmission is the language itself, as opposed to utterance-based models of cultural transmission (e.g. Croft, 2000) that make the learning task much more manageable.

---

[1] https://github.com/SonyCSLParis/Naming-Game

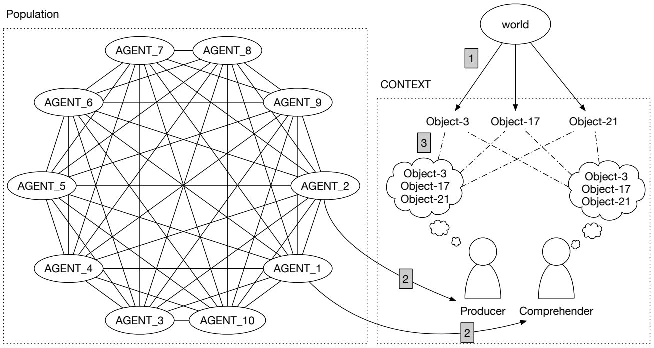


Figure 1. In the Language Game paradigm, two agents are drawn from a population of peers in order to engage in a locally situated communicative interaction.

## 3. The Naming Game

Let us now turn to the Language Game methodology, which is illustrated in Figure 1. A language game experiment typically involves a population of multiple agents, organized according to a network topology that represents the population's social structure. At each *time step* of the experiment, two agents are drawn from the population to play a *language game* with each other, which is a locally situated interaction that is private to the participating agents. Since all agents are peers, each agent can take on the role of producer or comprehender. During a language game, agents only worry about achieving their communicative goals with respect to the current situation, so they are not preoccupied by learning "the" community language. In fact, as will be demonstrated below, the community language spontaneously emerges and evolves through *self-organization* in very much the same way an ant path or other complex systems are formed in nature.

The simplest Language Game experiment is the Naming Game, which was first introduced by Steels (1995) and which has well-understood mathematical properties (Baronchelli et al., 2006; De Vylder & Tuyls, 2006). The Naming Game involves a population of $N$ agents that need to self-organize a shared lexicon $\mathcal{L}$ for referring to a number $M$ of objects present in their world. Each agent is endowed with an associative lexicon $L_a$ that consists of a list of *lexical constructions*, here operationalized as signal-referent associations (in order to be consistent with the example of Woensdregt et al., 2021) that are assigned a *preference score* that represents the strength of an association. More formally: $L_a = \{\langle s_1, r_1, \sigma_1 \rangle, ...\}$ where $s_i$ is a possible signal (a string), $r_i$ a possible referent (a unique symbol), and where $0.0 \leq \sigma_i \leq 1.0$ is the preference score with as initial value $0.5$. All agents start with an empty lexicon at time step $t = 0$, but gradually invent and learn new constructions as they interact with each other according to the following scenario:

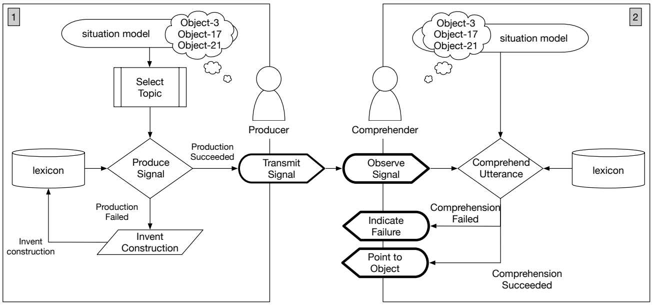


Figure 2. This Figure shows the mental operations that the producer goes through in order to produce an utterance, and the operations that the comprehender performs to comprehend the utterance.

1. Each time step $t$, a *situated context* is initialized, as illustrated in Figure 1.
   (a) A context consists of $n$ objects randomly selected from the "world."
   (b) One agent is randomly selected to act as the *producer*, and a *comprehender* is selected among the agents that are directly connected to the producer's social network (shown on the left of Figure 1).
2. The agents start interacting. Figure 2 illustrates their linguistic actions:
   (a) Both agents maintain a *situation model* of the current context, in which they keep track of the objects they perceive.
   (b) The producer randomly selects one of the objects as the *topic*.
   (c) The producer transmits a signal for referring to the topic to the comprehender:
      - If the producer does not know an appropriate signal yet for the topic, they will invent one. Here, they will randomly generate a string according to the template "CVCVCVCVCV" (e.g. "kebekobola");
      - If the producer knows more than one signal that associated with the topic ("competitors"), they will choose one according to the *inventory dynamics* of the experiment. In this paper, the agent chooses the construction that has the highest preference score.
   (a) The comprehender tries to comprehend the producer's signal.
      - If the comprehender knows a construction that maps the signal onto one of the referents in their situation model, comprehension succeeded. They will point to that object.
      - If not, the comprehender will indicate failure to the producer.

3. The producer gives feedback to the comprehender. They will signal success if the comprehender pointed to the correct object, and signal failure otherwise. In the latter case they also point to the intended topic, so the the comprehender can *learn* a new signal-referent association.
4. The agents update their linguistic inventories based on the success or failure of the game according to the experiment's inventory dynamics.

As can be inferred from the above, the *inventory dynamics* of an experiment determine how the agents cope with variation in the population and how they update their linguistic inventories after each usage event (see Baronchelli, 2018 for a primer on suitable strategies for achieving consensus). In the experiments of this paper, agents update their inventories using *lateral inhibition*, which means that they will increase the preference scores of lexical constructions that led to success while at the same time punishing competitors by lowering their preference scores, which is compatible with more recent proposals on statistical preemption in constructivist language learning (Goldberg, 2011). Agents will also punish constructions that led to communicative failure. The experiments in this paper adopt the score updating rule of De Beule, De Vylder, and Belpaeme (2006), which has been proven successful in prior research.

## 4. Experimental Results and Discussion

To remain close to the example of Woensdregt et al. (2021), a Naming Game experiment was set up with a population size of $N = 10$ and a number of objects $M = 25$. In order to test whether the population succeeds at self-organizing a lexicon for referring to these 25 objects, 100 independent simulations were executed with each 6.000 time steps, which amounts to an average of 1.200 interactions per agent per simulation.

Figure 3 shows the most important results averaged over the hundred simulations, with error bars indicating the variability between each run. The measure Communicative Success is a running average of the past 10 interactions in which a failed game counts as 0 and a successful game as 1. As can be seen, the agents already reach success after about 2.000 time steps (about 400 interactions per agent or 16 interactions per object). Communicative success however doesn't mean that agents have reached *consensus* about which signal to use for which referent, because each agent might simply have learned all of the signals in order to understand the others, but keep using their own preferences. The consensus of a signal-referent association is measured as the inverse of normalized (Shannon) entropy, as formally defined in the supporting code. Global consensus simply averages over these individual consensus scores. As can be seen in the left graph of Figure 3, consensus quickly follows communicative success, with maximal consensus after about 4.000 interactions, an average of 800 interactions per agent, or 32 interactions per referent.

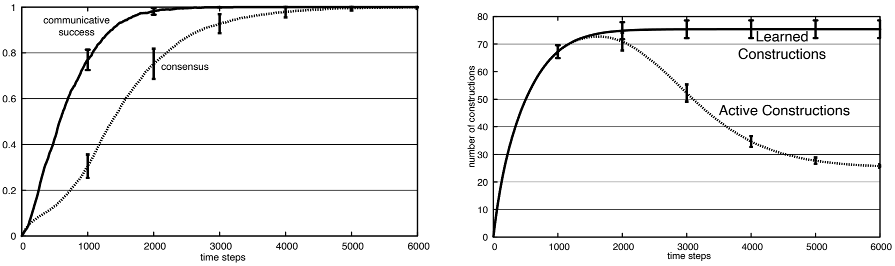


Figure 3. Naming Game with $N = 10$ and $M = 25$. Results average over 100 simulations. Left: Agents reach persistent communicative success after about 2.000 time steps and a global consensus after about 5.000 time steps. Right: Agents have to learn on average 75 words, but their alignment strategy allows them to reduce their active lexicon to an optimal size of one signal per referent.

The graph on the right shows more information about the emergent lexicon. Each agent learns about 75 different constructions for referring to 25 objects. However, through lateral inhibition the agents succeed in reducing their lexicons to an optimal size of 25 active constructions. Constructions are considered to be active as long as their preference score exceeds a threshold $\sigma_i \geq 0.2$, otherwise they become dormant. Hidden from the graph is the actual variation in the population: every time a producer communicates about an object for the first time, they invent a new construction. Given that the producer is randomly selected, on average half of the population $N/2$ will invent a competitor for the same object, leading to an average of five competitors per referent or a total of 125 words for 25 referents circulating in the population. If the agents would have to consider the space of all possible languages, as in the formalization of Woensdregt et al. (2021), they would have to entertain $2^{125\times 25}$, or about $5.2 \times 10^{939}$, possible languages. Yet, as readers who test the code can verify, it takes only a couple of seconds to run the simulations on a present-day laptop (in fact, it takes more time to produce the graphs than running the simulations themselves).

## 5. Discussion and Conclusion

The previous sections explained a simple Naming Game, but how does it scale? Prior research has provided mathematical proof that the Naming Game always converges (De Vylder & Tuyls, 2006), and Baronchelli et al. (2006) and Baronchelli (2006) have examined how the model behaves when scaling the population size, reporting simulations up to $N = 100000$. These experiments reveal that the model displays similar behavior as natural language dynamics, most notably that there is an S-shaped curve with sharp transition towards population-wide convergence. In fact, the transition becomes sharper and sharper as the size of the language increases. Baronchelli et al. (2006) conclude that this “surprising result [...] explains why human language can scale up to very large populations.”

Other experiments have shown how the Language Game method can also successfully apply to more complex languages where there is no one-to-one mapping between signal and referent, including large lexicons and meaning spaces (Wellens, Loetzsch, & Steels, 2008), or experiments on grammatical structures such as argument structure constructions (see Steels, 2012 for a collection of experiments, and the open-access book series *Computational Models of Language Evolution* at Language Science Press).

Why doesn't the Language Game method hit the wall of computational intractability? The answer is that the agents never exhaustively search the hypothesis space, but instead only consider what is *relevant* for achieving communicative success in their local interactions. For example, when a learner observes a particular signal, they will not try to update their entire lexicon, but only those constructions that were involved in the language game: the constructions that were used and their competitors. In the current setup, an agent knows on average three competitors for one referent, so they will on average never make more than three local adjustments (i.e. update their preference scores) at each time step instead of recalculating the probabilities of all possible mappings. Global consensus is nevertheless achieved as a side-effect in the same way as ant paths are spontaneously formed as the side-effect of local behaviour. Language can therefore be seen as a *complex adaptive system* that is constantly shaped and reshaped through language usage (Steels, 2000).

In sum, when Woensdregt et al. (2021) posed the challenge of scaling, they wrongly equated "models of language evolution" with a particular kind of model in which a language must be learned as a whole. Other models that operationalize language learning as a much more manageable task, such as utterance-based models of language evolution (Croft, 2000) or the Language Game paradigm, do not run into the problem of computational intractability; and have in the latter case already been demonstrated to scale to more realistic settings. And just like the Bayesian iterated learning models, the Language Game method is a general framework that is agnostic to the specifics of implementation, as can be gleaned from the breadth of techniques and phenomena that have already been investigated with this method (Steels, 2012).

## Acknowledgements

The work reported in this paper has been funded by the Sony Computer Science Laboratories Paris. The author wishes to thank Luc Steels for his invaluable feedback, as well as Martina Galletti and Inès Blin, and the many researchers who contributed through discussions and collaborations on the Babel software suite, particularly Katrien Beuls (Université de Namur) and Paul Van Eecke (VUB – KULeuven). Special thanks also go to Hiroaki Kitano, Vittorio Loreto and Peter Hanappe for creating such a superb work environment.

## References

Baronchelli, A. (2006). *Statistical mechanics approach to language games.* Unpublished doctoral dissertation, Università di Roma “La Sapienza”, Rome.

Baronchelli, A. (2018). The emergence of consensus: A primer. *Royal Society Open Science*, *5*(172189).

Baronchelli, A., Felici, M., Loreto, V., Caglioti, E., & Steels, L. (2006). Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics*, *P06014*.

Bybee, J. (2006). From usage to grammar: The mind’s response to repetition. *Language*, *82*(4), 711–733.

Croft, W. (2000). *Explaining language change: An evolutionary approach.* Harlow, Essex: Longman.

De Beule, J., De Vylder, B., & Belpaeme, T. (2006). A cross-situational learning algorithm for damping homonymy in the guessing game. In L. M. Rocha, L. S. Yaeger, M. A. Bedau, D. Floreano, R. L. Goldstone, & A. Vespignani (Eds.), *Artificial life x* (pp. 466–472). MIT Press.

De Vylder, B., & Tuyls, K. (2006). How to reach linguistic consensus: A proof of convergence for the naming game. *Journal of Theoretical Biology*, *242*(4), 818–831.

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, *22*(1), 131–154.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*, 441–480.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, *77*(2), 187–222.

Loetzsch, M., Wellens, P., De Beule, J., Bleys, J., & van Trijp, R. (2008). *The babel2 manual* (Tech. Rep. No. AI-Memo 01-08). Brussels: AI-Lab VUB.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Cambridge, MA: Harvard University Press.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, *2*(3), 319–332.

Steels, L. (2000). Language as a complex adaptive system. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. Merelo, & H.-P. Schwefel (Eds.), *Parallel problem solving from nature* (pp. 17–28). Berlin: Springer-Verlag.

Steels, L. (Ed.). (2012). *Experiments in cultural language evolution.* Amsterdam: John Benjamins.

Wellens, P., Loetzsch, M., & Steels, L. (2008). Flexible word meaning in embodied agents. *Connection Science*, *20*(2–3), 173–191.

Woensdregt, M., Spike, M., Haan, R. de, Wareham, T., Rooij, I. van, & Blokpoel, M. (2021). *Why is scaling up models of language evolution hard?* PsyArXiv. (https://doi.org/10.31234/osf.io/d2h5c, version of May 2021)

# EVO-DEVO FRAMEWORK: CONTEXTUALIZING LANGUAGE ONTOGENY IN LANGUAGE EVOLUTION

OLGA VASILEVA*[1] and ANTONIO BENITEZ-BURRACO[2]
*Corresponding Author: ovasileva@langara.ca
[1] Psychology Department, Langara College, Vancouver, Canada
[2]Linguistics Department, University of Seville, Seville, Spain

Language development in human infants has long been of interest to the field of language evolution. Topics of investigation include (but are not limited to) arguments regarding “language fossils” and potential parallels between ontogeny and phylogeny of language development (e.g. Bickerton, 1984; Wray, 1998), co``mparative research with language-trained primates (Savage-Rumbaugh et al., 1993) and arguments regarding language nativism (Berwick, Chomsky & Piatelli-Palmarini, 2013). Perhaps surprisingly, despite such a long-standing interest of the field, to this day there is no systemic approach to the question of language ontogeny in language evolution. This seems crucial since evolution depends on changes in ontogenetic patterns. Given advancements of relevant disciplines (evolutionary, developmental and systems biology) we argue that the time is ripe for developing such an approach. Specifically, we propose that translating the achievements of the Evolutionary Developmental Biology (Evo-Devo) to the field of language evolution is a particularly promising avenue for future research. Evo-Devo approach stresses that *development* is central to evolutionarily processes by becoming the source of phenotypic plasticity on which selection operates and optimizing trait alignment with the adaptive landscape (Jamnickzky et al., 2010; Michel et al., 2018). Moreover, due to similarities in molecular foundations of gene networks expression across distant species (i.e. *deep homology*), analogous or divergent pattern of the same developmental program unfolding in different species and even taxa can be informative for understanding specific selection pressures and convergent and divergent evolution, respectively (Fitch, 2012). Thus, the premise of Evo-Devo approach unites ontogeny and phylogeny on several levels going beyond simplistic claims of recapitulation and in the context of language emergence beyond simplistic parallels between processes of language acquisition by a child and its emergence in phylogeny. Ultimately, Eco-Evo-Devo theories support the view that organisms evolve as a result of the interactions between their genes, their developmental paths, and the environments in which they live. In the context of human language evolution this Eco-Evo-Devo framework opens the door to views of language evolution that see the origins of modern language as the result of both a biological process that changed our brain and behavior, and a cultural process that enabled the emergence of distinctive aspects of human languages through transmission and use.

In the talk, we will outline how application of this framework provides insights in the field of language evolution. First, in line with Evo-Devo approaches, we will advocate for a systemic approach to language acquisition where language is treated as a multicomponent system (Fitch, 2012). Instead of trying to single out *the component* of language that provides major evolutionary novelty either on the internal (syntax) or external (vocal learning) levels, we will support the view that it is more productive and accurate to investigate development of *systems*, both cognitive and behavioural, along with their interaction with the environment. This is in line with current comparative research, that is no more interested in finding *the trait* that could explain language, but how language results from a unique combination of many shared abilities (Fitch, 2020; Vasileva, 2019). Second, we will argue that since systems develop over the lifespan, constantly being under selection due to changing adaptive resources of an organism (Bateson 2014), the focus should be thus shifted towards the analysis of the emergence and interaction of varied components (e.g. vocabulary and ToM) over the life span, without limiting on early ontogeny. Finally, because development can be conceptualized as continuous bidirectional interactions between different levels from the molecular-DNA to the behavior - psychological ones, leading to trait variability and stability (Lickliter & Honeycutt, 2013; West-Ebenard, 2003), we will defend the view that it is key to investigate systemic patterns of variation in both typical and abnormal cognitive-linguistic development (Barceló-Coblijn, Benitez-Burraco & Irurtzun 2015), with a special emphasis on the *variation in developmental patterns*, rather than capturing trait association at a single time-point become crucial (Michel et al., 2018). Four major processes - heterochrony, heterotypy, heterometry and heterotopy - described in the Evo-Devo that can be applied to the study of language ontogeny to contextualize language ontogeny in language evolution and deserve a close examination.

## References

Barceló-Coblijn, LI., Benítez-Burraco, A., & Irurtzun, A. (2015). Syntactic networks as an endophenotype of developmental language disorders: An evo-devo approach to clinical linguistics. Biolinguistics, 9, 043-049.

Bateson, P. (2014). New thinking about biological evolution. Biological Journal of the Linnean Society, 112(2), 268-275.

Berwick, R. C., Chomsky, N., & Piattelli-Palmarini, M. (2013). Poverty of the stimulus stands: Why recent challenges fail. Rich languages from poor inputs, 19-42.

Bickerton, D. (1984). The language bioprogram hypothesis. Behavioral and brain sciences, 7(2), 173-188.

Fitch, W. T. (2012). Evolutionary developmental biology and human language evolution: constraints on adaptation. Evolutionary biology, 39(4), 613-637.

Fitch, W. T. (2020). Animal cognition and the evolution of human language: why we cannot focus solely on communication. Philosophical Transactions of the Royal Society B, 375(1789), 20190046.

Gottlieb, G. (2000). Environmental and behavioral influences on gene activity. Current Directions in Psychological Science, 9, 93–102. https://doi.org/10.1111/1467-8721.00068

Jamniczky, H. A., Boughner, J. C., Rolian, C., Gonzalez, P. N., Powell, C. D., Schmidt, E. J., ... & Hallgrímsson, B. (2010). Rediscovering Waddington in the post-genomic age: Operationalising Waddington's epigenetics reveals new ways to investigate the generation and modulation of phenotypic variation. Bioessays, 32(7), 553-558.

Lickliter, R., & Honeycutt, H. (2013). A developmental evolutionary framework for psychology. Review of General Psychology, 17(2), 184-189.

Michel, G. F., Babik, I., Nelson, E. L., Campbell, J. M., & Marcinowski, E. C. (2018). Evolution and development of handedness: An Evo–Devo approach. Progress in brain research, 238, 347-374.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., & Bates, E. (1993). Language comprehension in ape and child. Monographs of the society for research in child development, i-252.

Vasileva, O. (2019). Beyond “uniqueness”: Habitual traits in the context of cognitive-communicative continuity. Theoria et Historia Scientiarum, 16, 129.

West-Eberhard, M.J., 2003. Developmental plasticity and evolution. Oxford University Press, Oxford

Wray, A. (1998). Protolanguage as a holistic system for social interaction. Language & communication, 18(1), 47-67.

# LANGUAGE DEVELOPMENT AND COGNITIVE ONTOGENY: AN IMPORTANT LINK

OLGA VASILEVA*[1] and MICHAEL PLEYER[2]

*Corresponding Author: ovasileva@langara.ca
[1]Psychology Department, Langara College, Vancouver, Canada
[2]CLES, Nicolaus Copernicus University, Torun, Poland

The field of language of evolution is inherently interdisciplinary, attracting specialists from varied disciplines spanning from linguistics and psychology to computer science and biology. Although this interdisciplinary nature is incredibly productive and should be celebrated, it poses additional challenges to the field. Among such challenges is the existence of somewhat neglected topics that are crucial for understanding language evolution and yet are not sufficiently investigated, such as, for example, animal cognition (Fitch, 2020). Recent advancements in the field call for a more holistic approach to language by treating it first and foremost as a multi component system (Benítez-Burraco & Boeckx, 2014; Fitch, 2012; 2020) that is intimately related to human cognition. It is this tight link between organisms' communicative system and cognitive abilities that is unmatched in other species and makes human language particularly "stand out". Although the very idea of the role of language in the formation of human cognition is not novel (tracing back to Vygotsky (2012), it has received surprisingly little attention from the field of language evolution.

Language ontogeny is frequently invoked in language evolution discussions. At the same time, it is a research area that highlights the sharp contrast between two approaches focusing on the *communicative* and *"purely linguistic"* aspects of language. While the former investigate communicative aspects of language and other parameters that likely interact with the language system, such as social cognition (Tomasello at al., 2005), the latter focuses on the development of the linguistic system (e.g. syntax or vocabulary). It is thus possible to say that these approaches investigate either the formation of the linguistic system and its internal "gear" (e.g. syntax) or language precursors or the external aspects of language (e.g. speech production), with insufficient interaction between the two. This situation in turn results in a quite fragmented (if not reductionist) approach where child development is studied not as a unified phenomenon, but rather specific aspects of it at a time point. An Evolutionary-Developmental Linguistics remains quite underdeveloped as a research field.

In our opinion, language development provides one of the most convincing types of evidence of the connection between language and cognition in humans. We briefly outline two lines of research demonstrating that in human children language development 1) corresponds with simultaneous and inter-dependent emergence of varied cognitive abilities and 2) results in qualitative changes in behavior. From the first line of research examples of language association with other domains include: object permanence (Gopnik & Meltzoff, 2021), categorization (Novack et al., 2021), and ToM (Tomasello et al., 2005). Research on the interaction of the motor domain and language suggests that developing linguistic system allows children to perform more complex hierarchical goal-oriented behaviors (Greenfield, 1991; Michel et al., 2013) as well as engage in complex play behavior (Pleyer, 2020).

Summarized results of decades long research programs with primates trained to communicate with language-like systems suggests that primates are overall capable of using symbolic communicative systems (Savage-Rumbaugh et al., 2018). Additionally, results of comparative studies demonstrate the presence of cognitive abilities in other animals remarkable both in their complexity and abundance (Andrews, 2020; Zentall, 2020). Animals can represent their environment mentally, operate abstract concepts and perform goal-oriented behavior (Godfrey-Smith, 2016). However, to this day, there is no convincing evidence the development of the communicative system has drastic effects on the development of cognitive abilities and qualitative changes in the behavior in other species (Novack & Waxman, 2020). While absence of evidence, by all means, cannot be equated with evidence of absence, this notion requires careful theoretical and empirical investigation. Underestimation of the importance of the link between human cognition and language in ontogeny in our opinion leads to limitations in developing language evolution research programs. We suggest that it is important not only to compare and contrast animal communication and communicative functions in human language but to explore differences in cognitive ontogeny as well as pose the question of why and how the tight link between language and cognition emerged in humans.

## References

Andrews, K. (2020). The animal mind: An introduction to the philosophy of animal cognition. Routledge.

Benítez-Burraco, A., & Boeckx, C. (2014). Universal Grammar and biological variation: an EvoDevo agenda for comparative biolinguistics. Biological Theory, 9(2), 122-134.

Godfrey-Smith, P. (2016). Other minds: The octopus and the evolution of intelligent life. London: William Collins.

Gopnik, A., & Meltzoff, A. N. (2021). Early semantic developments and their relationship to object permanence, means-ends understanding, and categorization. In Children's language (pp. 191-212). Psychology Press.

Fitch, W. T. (2012). Evolutionary developmental biology and human language evolution: constraints on adaptation. Evolutionary biology, 39(4), 613-637.

Fitch, W. T. (2020). Animal cognition and the evolution of human language: why we cannot focus solely on communication. Philosophical Transactions of the Royal Society B, 375(1789), 20190046

Novack, M. A., Brentari, D., Goldin-Meadow, S., & Waxman, S. (2021). Sign language, like spoken language, promotes object categorization in young hearing infants. Cognition, 215, 104845.

Novack, M. A., & Waxman, S. (2020). Becoming human: human infants link language and cognition, but what about the other great apes?. Philosophical Transactions of the Royal Society B, 375(1789), 20180408.

Pleyer, M. (2020). Pretend play as a scaffold for language evolution. In A. Ravignani, C. Barbieri, M. Martins, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, K. Mudd, & T. Verhoef (Eds.). The evolution of language: Proceedings of the 13th international conference (EvoLang13). https://doi.org/10.17617/2.3190925

Savage-Rumbaugh, S., Roffman, I., Lingomo, S., & Pugh, E. (2018). The fully conscious ape. International Journal of Comparative Psychology.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. Behavioral and brain sciences, 28(5), 675-691.

Vygotsky, L. S. (2012). Thought and language. MIT press.

Zentall, T. R. (2020). Animal intelligence. In R. J. Sternberg (Ed.), The Cambridge handbook of intelligence (pp. 397–427). Cambridge University Press.

# LANGUAGE EVOLUTION WITHOUT TRUST

TILL NIKOLAUS VON HEISELER*[1], ANNE AERTS[14], KAI SHAMAN[2], CHRISTOPHER DROPPA[2], DAPHNE PRESTON-KENDAL[3], FRANZISKA EBERHARDT[5], MAURICE PFLUG[2], ANNA MAASER[3], CLAUDIO FABBRONI[2], VICTORIA CHARLOTT UHLICH[2]

*Corresponding Author: formatlabor.net@gmail.com
[1]Department of Philosophy, [2]Berlin School of Mind and Brain, [3]Department of German Studies and Linguistics, [4]Department of Social Sciences, [5]Institute for Biology, Humboldt University, Berlin, Germany

Most researchers in the field of language evolution share the assumption that *before* the evolution of language could even begin, a complex social development must have taken place, establishing new forms of collaboration and trust between the members of a group, not seen among any other apes. The reason for this is that linguistic communication requires trust since it is typically not backed up by direct evidence. According to Dor, Knight, and Lewis (2014, p. 4), virtually all researchers agree that an *unprecedented* level of collective cooperation and trust *was* needed for the first steps toward the evolution of language to occur. This paper aims to show that this dominant assumption is unnecessary.

We support our claim through a methodological suggestion: the investigation into language evolution should *not* begin with the traditional assumptions about language, but with the use of contemporary evolutionary theory to inform us how language should be conceptualized to be evolutionary explainable. One well-supported view is the *function-first approach* (von Heiseler, 2020), where a new function is initially fulfilled by a *behavioral shift* that exploits an already existing structure (Mayr, 2001, pp. 224-229). Since in a novel configuration, the new behavior is adaptive, the elements that were used for this behavior will adapt by natural selection to fulfill the new function. This process usually begins with a change in the *evolutionary configuration* (an arrangement of the physiological, environmental, and social components and constraints that jointly constitute the development and sustainment of an adaption by natural selection), making the new function adaptive.

We apply the *function-first approach* to the evolution of language by suggesting that the proper evolutionary function of language is to convey propositions, especially about *displaced* actions. There are two mutually exclusive and jointly exhaustive classes of signs: natural and non-natural signs (Grice, 1957). In a world without non-natural signs (such as mimetic or conventional symbols), only natural signs (if anything) can refer to a displaced action. An example of such referential displacement can be found in the display of objects that imply a past action of the communicator, e.g., communicating a past kill by displaying the hunted animal as a trophy. Such use of natural signs could be reproductively beneficial for an individual by enhancing their attractiveness to potential partners for cooperation, or mating, or social status in general.

The *function-first approach* predicts that language evolved from such display behavior. This proposal has the potential to solve both major problems of language evolution—the *problem of the graduality of the emergence of syntax* and the *bootstrapping problem of language and cooperativeness/trust*. The problem of graduality is solved by identifying a structural similarity between understanding the display of an indexical object implying a past action—such as a trophy—and the simplest linguistic syntactic patterns—that of the simple transitive sentence: the presenter of the trophy constitutes the *agent*, the trophy represents the *patient*, and the verb is implied by the state of the *patient* (von Heiseler, 2019). By integrating a single mimetic gesture that represents an action by reenactment and which marks the thematic roles by being directed from the *agent* to the *patient*, a simple proposition is expressed. Moreover, since this approach suggests that the first language-like communication included indexical objects as evidence, the problem of trust disappears. We show how the function of referring to displaced actions could have evolved further in various niches of trust, such as in mother-child communication and teaching (kin selection; Fitch, 2004, Gärdenfors & Högberg, 2022), displaying performative qualities (a quality in which individuals vary naturally; Mithen, 2005), and bonding (building reciprocal relations; Dunbar, 2011). Thus, as these two problems, which are intractable in the dominant view of language evolution, are solvable under our proposal, we suggest that the best course of action in language evolution research is to search for evolutionary configurations in which referential *displacement* and *propositional communication* became adaptive (see e.g., Bickerton & Szathmáry, 2011, von Heiseler, 2022). We conclude with the hope that we have shown that the hypothesis that an unprecedented level of collective cooperation was needed before the first step towards the evolution of language could have occurred, is not only unnecessary but is also unwarranted.

## References

Bickerton, D., & Szathmáry, E. (2011). Confrontational scavenging as a possible source for language and cooperation. *BMC Evolutionary Biology*.

Dor, D., Knight, C., & Lewis, J. (2014). Introduction. In D. Dor, C. Knight, & J. Lewis (Eds.), *The Social Origins of Language*. Oxford: Oxford University Press.

Dunbar, R. (2011). Gossip and the social origins of language. In M.Tallerman, & S.Gibson (Eds.), *Oxford Handbook of Language Evolution.* Oxford: Oxford University Press.

Fitch, W. T. (2004). Kin selection and mother tongues: a neglected component in language evolution. In U. Griebel, & D. K. Oller (Eds.), *Evolution of communication systems: a comparative approach* (pp. 275–296). Cambridge, Massachusetts: MIT Press.

Gärdenfors, P., & Högberg, A. (2022). Evolution of intentional teaching. In N. Gontier, A. Lock, & C. Sinha, *The Oxford Handbook of Human Symbolic Evolution.* Oxford.

Grice, P. H. (1957). Meaning. *The Philosophical Review, 66*, pp. 377 - 388.

Mayr, E. (2001). *What Evolution Is.* New York: Basic Books.

Mithen, S. (2005). *The Singing Neanderthals: the origins of music, language.* Harvard University Press.

von Heiseler, T. N. (2019). Syntax of Testimony: Indexical Objects, Syntax, and Language or How to tell a story without words. *Frontier in Psychology*.

von Heiseler, T. N. (2020). The Social Origin of the Concept of Truth – How Statements Are Built on Disagreements. *Frontiers in Psychology*.

von Heiseler, T. N. (2022). How Language and Human Altruism Evolved Hand in Hand — The Backchannel Hypothesis. *Frontiers in Psychology*.

# UNPACKING THE PROXIMATE FUNCTIONS OF FUNCTIONALLY REFERENTIAL SIGNALS: A TEST OF 3 HYPOTHESES

Stuart K Watson*[1,2,3], Andrea Müller[3], and Marta Manser[2,3]

*Corresponding Author: swatso88@gmail.com

[1] Department of Comparative Language Science, University of Zurich, Switzerland

[2] Center for the Interdisciplinary Study of Language Evolution, University of Zurich, Switzerland

[3] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

Decades of research have demonstrated a broad range of species possess vocal signals which are produced only in the presence of specific stimuli, such as a particular predator type. Crucially, conspecific receivers are able to decode these stimulus-specific associations and use them to choose an appropriate behavioural response to the eliciting cause (Arnold & Zuberbühler, 2013; Dawson Pell et al., 2018; R. Seyfarth et al., 1980). Because of the surface-level resemblance of these call-types to referential language, these have been termed “functionally referential” vocalisations, and have been argued to present a likely evolutionary precursor to semantics (R. M. Seyfarth & Cheney, 2017). However, plausible competing explanations for the underlying function of these calls have also been presented. For example, stimulus-specific calls may be a reflexive expression of the internal arousal state brought about by a particular stimulus, which receivers have simply learned to associate with the eliciting cause (Wheeler & Fischer, 2012). Alternatively, they may be more akin to ‘imperative’ statements in language, functioning primarily to direct a specific behavioural change in receivers (Schamberg et al., 2018). Determining the proximate function of these signals is therefore key to understanding both animal communication systems and the evolutionary foundations of language. Unfortunately, the underlying function of these calls from a producer perspective is not trivial to determine. Playback experiments are a powerful tool for revealing how receivers interpret and extract information from a given call, but cannot directly inform us about the motivation of callers themselves. Here, we use a novel model predator demonstration experiment to shed light on the proximate function of aerial-specific alarm call production in meerkats (*Suricata suricatta*), a cooperatively breeding species of mongoose with a complex communication system (Manser et al., 2014).

Meerkats have an expansive vocal repertoire, one item of which is a functionally referential alarm call produced only in response to aerial predators (typically martial eagles) (Manser et al., 2002). In the Kalahari Desert, we simulated predation events for 7 groups of wild meerkats (4-22 individuals per group) using a model predator, a kite painted to imitate a martial eagle. Demonstrations took place in two behavioural contexts: i) while the majority of the group was safe and close (<5m) to their burrows (N = 3 trials per group), and ii) while the majority of the group was vulnerable, foraging at a distance (>10m) from their burrows (N = 3 trials per group). All trials elicited strong predator-avoidance response behaviours from the meerkats. We recorded all calling bouts produced in response to these predator demonstrations and coded whether they contained an aerial alarm call. Contextual information was also recorded, including: the distance of the caller to shelter, the proportion of the group who were close to shelter, and the distance of the kite from the group. The association between contextual variables and the production of aerial alarm calls was then used to test three hypotheses regarding the underlying function of these calls (Table 1).

**Table 1.** Summary of hypotheses for the underlying function of meerkat aerial alarm calls and their corresponding empirical predictions.

| **Hypothesis** | | **Predictions** |
|---|---|---|
| Referential | Produced to share contextual information with receivers | Production will not vary according to the relative danger towards the caller or group |
| Imperative | Produced to elicit a specific behaviour in receivers | Production is more likely when the group (i.e. receivers) are further from safety |
| Emotional | Produced reflexively as a product of arousal state | Production is more likely when the caller is further from safety / closer to danger |

To determine which of these three hypotheses best explained the observed data, we fit a suite of Bayesian generalised linear mixed-effects models with effect structures corresponding to each of these hypotheses. The out-of-sample predictive power of each of these models was then compared using Watanabe-Akaike information criterion. We found that the model which best fit the data was one in which no fixed effects corresponding to the relative safety of the caller or group influenced the likelihood of an individual producing an aerial alarm call, supporting the referential hypothesis.

This finding indicates that the production of aerial alarm calls in meerkats primarily serves a referential function, i.e., to inform receivers about the presence of a particular predator type in their immediate surroundings. This novel approach provides the first insights into the cognition and motivations underlying the production of these functionally referential signals Broader application of this approach to the functionally referential signals of other species will determine whether this is a general feature of this class of signal, or if there is variation within and between species.

## Acknowledgements

SKW and MM were funded by NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40_180888.

## References

Arnold, K., & Zuberbühler, K. (2013). Female Putty-Nosed Monkeys Use Experimentally Altered Contextual Information to Disambiguate the Cause of Male Alarm Calls. *PLOS ONE*, *8*(6), e65660. https://doi.org/10.1371/journal.pone.0065660

Dawson Pell, F. S. E., Potvin, D. A., Ratnayake, C. P., Fernández-Juricic, E., Magrath, R. D., & Radford, A. N. (2018). Birds orient their heads appropriately in response to functionally referential alarm calls of heterospecifics. *Animal Behaviour*, *140*, 109–118. https://doi.org/10.1016/j.anbehav.2018.04.010

Manser, M. B., Jansen, D. A. W. A. M., Graw, B., Hollén, L. I., Bousquet, C. A. H., Furrer, R. D., & le Roux, A. (2014). Vocal Complexity in Meerkats and Other Mongoose Species. In *Advances in the Study of Behavior* (Vol. 46, pp. 281–310). Elsevier. https://doi.org/10.1016/B978-0-12-800286-5.00006-7

Manser, M. B., Seyfarth, R. M., & Cheney, D. L. (2002). Suricate alarm calls signal predator class and urgency. *Trends in Cognitive Sciences*, *6*(2), 55–57. https://doi.org/10.1016/S1364-6613(00)01840-4

Schamberg, I., Wittig, R. M., & Crockford, C. (2018). Call type signals caller goal: A new take on ultimate and proximate influences in vocal production. *Biological Reviews*, *93*(4), 2071–2082. https://doi.org/10.1111/brv.12437

Seyfarth, R., Cheney, D., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803. https://doi.org/10.1126/science.7433999

Seyfarth, R. M., & Cheney, D. L. (2017). The origin of meaning in animal signals. *Animal Behaviour*, *124*, 339–346. https://doi.org/10.1016/j.anbehav.2016.05.020

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology: Issues, News, and Reviews*, *21*(5), 195–205. https://doi.org/10.1002/evan.21319

# USING MACHINE LEARNING TO SHOW CHIMPANZEE (PAN TROGLODYTES SCHWEINFURTHII) LIP-SMACK RHYTHM VARIATION: IMPLICATIONS FOR THE STUDY OF HUMAN SPEECH EVOLUTION.

CHARLOTTE WILTSHIRE*[1], ANDRÉ PEREIRA[2], GAL BADIHI[1], ADRIAN SOLDATI[1,3], KATIE SLOCOMBE[4], KLAUS ZUBERBUEHLER[1,3], ADRIANO REIS E LAMEIRA[5], and CATHERINE HOBAITER[1]

*Corresponding Author: cw266@st-andrews.ac.uk
[1] School of Psychology and Neuroscience, University of St Andrews, St Andrews, UK
[2] School of Psychology, University of Exeter, Exeter, UK
[3]School of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[4] School of Psychology, University of York, York, UK
[5] School of Psychology, University of Warwick, Coventry, UK

## 1. Introduction

Despite the incredible diversity of sound and structure, all human languages have an open-close mouth rhythm of 2-7Hz (Chandrasekaran et al., 2009) that is essential for speech perception (Shannon et al., 1995; Smith et al., 2002). Chimpanzee lip-smacks are an affiliative signal recently found to match the rhythm of human speech (Pereira et al., 2020) and, unlike other primate signals displaying this rhythm (Ghazanfar et al., 2012; Toyoda et al., 2017; Terleph et al., 2018), show substantial variation between and within populations. Human speech rhythm also varies within the 2-7Hz range between languages (Abercrombie, 1967; Ordin and Polyanskaya, 2015) and within them depending on the addressee (Wynn et al., 2018). The recent evidence that chimpanzee lip-smacks vary within the 2-7Hz band suggests that this important step-change in flexibility occurred before our split from other hominids. Whilst suggesting important similarities to human speech, it remains unclear whether lip-smack rhythm variation described (Periera et al. 2020) in chimpanzees was under individual control, and/or shaped by genetic or environmental factors. Here we further explore variation in chimpanzee lip smacks and assess the impact of features of individual (identity, age, rank), partner (identity, relative rank), and group identity. To increase sample size and improve measurement reliability, we employed a novel machine learning approach using DeepLabCut (Mathis and Mathis, 2020).

## 2. Methods

We recorded videos of lip-smacking from 2007-2020 in two communities within the Budongo population of East-African chimpanzees. We extracted 161 lip-smacking bouts from 26 individuals (19 males, 7 females, age 9-45 years) We created a bespoke chimpanzee facial tracker, to track 10 key-points using DeepLabCut (Mathis and Mathis, 2020). The effect of an individual's community, identity, age, sex, relative rank to their grooming partner or the sex of their grooming partner on bout rhythm was tested with a GLMM.

## 3. Results

Lip-smack rhythm varied from 1.52hz-9.49hz, with a mean frequency of 2.9hz (± 1.67Hz). The Waibira community showed a mean of 2.9hz (± 1.67Hz) and Sonso a mean of 2.87hz (±1.57Hz). We found no effect of the test variables or their interactions on rhythm. No two bouts displayed by the same individual showed the same rhythm, even within a grooming event with the same partner.

## 4. Discussion

We employed novel machine learning techniques to generate substantial sample sizes and reveal a new-found degree of flexibility in chimpanzee lip-smack rhythm. Rhythm falls within the range of speech, averaging towards the lower end of the range. This supports previous findings that primate rhythmic facial signals may be a precursor to speech (Morrill et al. 2012; Periera et al. 2020) rather than being dictated by other rhythmic mouth movements, such as chewing. This is further supported by substantial within-individual variation, which may have presented itself as within- and between-group variation in previous studies with small datasets (see: Periera et al. 2020).

Rhythm varied within a grooming bout, where multiple context factors are unlikely to have changed. Like speech, lip-smack rhythm is not fully explained by an individual's identity, sex, age, community, relative rank, or their partner's sex. Instead, they appear highly flexible, potentially reflecting context specific, rapidly changing features within a grooming bout. Controlling for context changes occurring during a grooming bout may offer further insight into what is driving variation. Within-bout variation in rhythm may suggest high levels of fine motor control, allowing chimpanzees to quickly change the timing of their lip-smacking during its production, as human's can to aid their communication during speech (Wynn et al., 2018).

## Acknowledgements

This research was funded by the European Research Council under grant agreement number 802719 and the St Andrews Restarting Research fund. We thank the field assistants and staff of the Budongo Conservation Field Station and the Greater Mahale Ecosystem Research Conservation station for their support. We thank the Ugandan Wildlife Authority, the Ugandan National Council for Science and Technology, the Tanzanian Wildlife Research Institute, and the Tanzanian Commission for Science and Technology for permissions to conduct research.

## References

Abercrombie, D. (1967). Elements Of General Phonetics. Edinburgh: Edinburgh University Press, pp. 112-146.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A.A. (2009). The natural statistics of audio-visual speech. PLoS Computational Biology, 5(7), e1000436, DOI: https://doi.org/10.1371/journal.pcbi.1000436.

Ghazanfar, A.A., Takahashi, D.Y., Mathur, N. and Fitch, T.W. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. Current biology, 22, pp. 1176-1182.

Mathis, M.W. and Mathis, A. (2020). Deep learning tools for the measurement of animal behavior in neuroscience. Current Opinion in Neurobiology, 60, pp. 1-11.

Morrill, R.J., Paukner, A., Ferrari, P.F. and Ghazanfar, A.A. (2012). Monkey Lip-smacking Develops Like the Human Speech Rhythm. Developmental Science, 15(4), pp. 557-568.

Ordin, M. and Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. The Journal of the Acoustical Society of America, 138(2), pp. 533-544.

Pereira, A.S., Kavanagh, E., Hobaiter, C., Slocombe, K.E. and Lameira, A.R. (2020). Chimpanzee lip-smacks confirm primate continuity for speech-rhythm evolution. Biology Letters, 16(5), 20200232, DOI: http://dx.doi.org/10.1098/rsbl.2020.0232.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M. (1995). Speech recognition with primarily temporal cues, Science, 270(5234), pp. 303-304.

Smith, Z.M., Delgutte, B., Oxenham, A.J. (2002). Chimaeric sounds reveal dichotomies in auditory perception, Nature, 416(6876), pp. 87-90.

Terleph, T.A., Malaivijitnond, S., Reichard, U.H. (2018). An analysis of white-handed gibbon male song reveals speech-like phrases. American Journal of Physical Anthropology, 166(3), pp. 649-660.

Toyoda, A., Maruhasi, T., Malaivijitnond, S. and Koda, H. (2017). Speech-like orofacial oscillations in stump-tailed macaque (*Macaca arctoides*) facial and vocal signals. American Journal of Physical Anthropology, 164(2), pp. 435-439.

Wynn, C.J., Borrie, S.A. and Sellers, T.P. (2018). Speech rate entrainment in children and adults with and without autism spectrum disorder. American Journal of Speech-Language Pathology, 27(3), pp. 965-974.

# EVOLUTIONARY DYNAMICS OF MULTILINGUAL COMPETITION UNDER INTERVENTION

Zhijun Wu

Department of Mathematics, Iowa State University, Ames, Iowa, USA
Email: zhijun@iastate.edu

An evolutionary dynamics model is investigated for the evolution of multilingual populations. The model consists of two different parts, formulated as two different evolutionary games. The first part accounts for the selection of languages based on the competition for popularity and social or economic advantages. The second part relates to the circumstance when the selection is altered, for better or worse, by forces other than competition such as public policies, education, or family influences. By combining competition with intervention, the model shows how a multilingual population may evolve under these two different sources of influences, and the languages may co-exist in evolutionarily stable multilingual forms with appropriate interventional measures. This is in contrast with the predictions from previous studies that the co-existence of languages is unstable in general, and one language will eventually dominate while all others become extinct.

## 1. Introduction

While many languages are in danger of extinction, multilingualism is being adopted as a common practice for communication among different language groups, and is playing a unique role in preserving language and cultural diversities (Gabszewicz et al., 2011; Grosjean, 2012). Languages compete and spread among their speakers, as genes are inherited and passed down to biological generations, where some are selected while others become extinct (Pinker & Bloom, 1990). Genes may be carried over in mixed forms. So are languages by multilingual speakers. In this paper, a mixed population of multilingual speakers and bilingual speakers in particular is considered, with multilingual defined broadly as zero, limited, or full uses of multiple languages or dialects, and an evolutionary dynamics model for its evolution is proposed, similar to that for genetic evolution (Burger, 2000).

Unlike genetic evolution though, the uses of languages are not only dependent of competition, but also subject to various societal interventions, common in social or cultural evolution. The proposed model consists of two different parts accordingly, formulated as two different evolutionary games, respectively. The first part accounts for the selection of languages based on the competition for popularity and social or economic advantages. The second part relates to the circumstance when the selection of languages is altered, for better or worse, by forces other than competition such as public policies, education, or family influences.

Much work has been done on modeling language competition, although not specifically for the evolution of multilingualism. A well known model was proposed by Abrams & Strogatz, 2003 for the study of language death. The model was later extended to more general and complex cases by several other groups (Mira & Paredes, 2005; Patriaca & Heinsalu, 2009; Vazquez et al., 2010; Fujie et al., 2013). The models along this line focus mainly on language competition for popularity and social or economic advantages, and predict that one language will eventually dominate while all others become extinct, and the co-existence of languages is unstable and hard to sustain.

While successfully applied to some language populations, the previous models have not explicitly distinguished language competition from possible societal interventions that may reverse the course of language changes. By combining language competition with possible societal interventions, this paper shows how a multilingual population may evolve under these two different sources of influences. It shows in particular that the co-existence of languages can be made stable and language extinction can be prevented with appropriate interventional measures, as seen in many multilingual communities across the world (Shin & Kominski, 2010; Grosjean, 2012; Batalova & Zong, 2016; China, 2017).

## 2. The Evolutionary Dynamics Model

Consider a bilingual population, the simplest yet the most common multilingual population. Assume it is large and well mixed, i.e., every individual speaker can interact with all others in the population. If an individual speaker uses two languages $A$ and $B$ with frequencies $x_A$ and $x_B$, respectively, this individual is called an $(x_A, x_B)$-speaker, where $0 \leq x_A, x_B \leq 1$, and $x_A + x_B = 1$. Likewise, if language $A$ and $B$ are used with frequencies $y_A$ and $y_B$ in average in the whole population, this population is called a $(y_A, y_B)$-population, where $0 \leq y_A, y_B \leq 1$, and $y_A + y_B = 1$.

First, consider a competition-only population. Let $P_A(y_A)$ and $P_B(y_B)$ be the payoff functions for $A$ and $B$ speakers in a $(y_A, y_B)$-population, respectively, defined in terms of the population sizes $y_A$ and $y_B$ and some other parameters for social or economic impacts, with $P_A$ increasing in $y_A$ and $P_B$ in $y_B$, meaning that the larger the population size of a language, the more benefit the language provides for its speakers. Then, the payoff function for an $(x_A, x_B)$-speaker in a $(y_A, y_B)$-population can be defined in terms of the average use of $A$ and $B$ by this speaker:

$$\pi((x_A, x_B), (y_A, y_B)) = x_A P_A(y_A) + x_B P_B(y_B). \tag{1}$$

Now consider the situation where the use of languages is influenced by some societal decisions. Assume that the societal interventions are implemented to counter the arbitrary increase or decrease of either language. Let $\bar{P}_A(y_A)$ and $\bar{P}_B(y_B)$ be the payoff functions for $A$ and $B$ speakers in a $(y_A, y_B)$-population, respectively, defined in terms of the population sizes $y_A$ and $y_B$ and some parameters for language reversing, with $\bar{P}_A$ decreasing in $y_A$ and $\bar{P}_B$ in $y_B$, meaning that the smaller the population size of a language, the more incentive or less penalty for the speakers of the language. Then, the payoff function for an $(x_A, x_B)$-speaker in a $(y_A, y_B)$-population can be defined in terms of the average use of $A$ and $B$ by this speaker:

$$\bar{\pi}((x_A, x_B), (y_A, y_B)) = x_A\bar{P}_A(y_A) + x_B\bar{P}_B(y_B). \tag{2}$$

The above two types of payoff functions can be combined to obtain an evolutionary dynamics model for bilingual competition under societal intervention:

$$\begin{cases} \dot{y}_A = y_A y_B(\tilde{P}_A(y_A) - \tilde{P}_B(y_B)) \\ \dot{y}_B = y_B y_A(\tilde{P}_B(y_B) - \tilde{P}_A(y_A)), \end{cases} \tag{3}$$

where

$$\tilde{P}_A(y_A) = \lambda P_A(y_A) + (1-\lambda)\bar{P}_A(y_A), \tag{4}$$

$$\tilde{P}_B(y_B) = \lambda P_B(y_B) + (1-\lambda)\bar{P}_B(y_B). \tag{5}$$

where $0 \le \lambda \le 1$. The model is reduced to competition-only when $\lambda = 1$, and to intervention-only when $\lambda = 0$.

Based on evolutionary game theory (Weibull, 1995; Hofbauer & Sigmund, 1998), the equations in (3) form a so-called system of replicator equations, which corresponds to an evolutionary game, with the Nash equilibrium being a strategy $(x_A^*, x_B^*)$ such that

$$\tilde{\pi}((x_A^*, x_B^*), (x_A^*, x_B^*)) \ge \tilde{\pi}((x_A, x_B), (x_A^*, x_B^*)) \text{ for all } (x_A, x_B), \tag{6}$$

where $\tilde{\pi}$ is the payoff function for the game, and for an $(x_A, x_B)$-speaker in a $(y_A, y_B)$-population

$$\begin{aligned} &\tilde{\pi}((x_A, x_B), (y_A, y_B)) \\ &= \lambda\pi((x_A, x_B), (y_A, y_B)) + (1-\lambda)\bar{\pi}((x_A, x_B), (y_A, y_B)). \end{aligned} \tag{7}$$

In general, $P_A$ and $P_B$ can be some increasing functions and $\bar{P}_A$ and $\bar{P}_B$ be some decreasing functions. However, in this study, they are defined using the following empirical functions similar to those in Abrams & Strogatz, 2003:

$$P_A(y_A) = c y_A^{\alpha-1} s_A, \ \ P_B(y_B) = c y_B^{\alpha-1} s_B, \ 1 < \alpha \le 2, \tag{8}$$

$$\bar{P}_A(y_A) = \bar{c} y_A^{\bar{\alpha}-1} \bar{s}_A, \ \ \bar{P}_B(y_B) = \bar{c} y_B^{\bar{\alpha}-1} \bar{s}_B, \ \ 0 \le \bar{\alpha} < 1, \tag{9}$$

where $c$ and $\bar{c}$ are scaling constants, $\alpha$, $\bar{\alpha}$, $s_A$, $s_B$, $\bar{s}_A$, $\bar{s}_B$ are all parameters, $0 \leq s_A, s_B, \bar{s}_A, \bar{s}_B \leq 1$. The parameters $\alpha$, $\bar{\alpha}$ determine the order of dependency of the payoffs on the population sizes. Since $1 < \alpha \leq 2$, the payoffs from $P_A$ and $P_B$ increase with increasing population sizes. On the other hand, since $0 \leq \bar{\alpha} < 1$, the payoffs from $\bar{P}_A$ and $\bar{P}_B$ decrease with increasing population sizes. The parameters $s_A, s_B$ are used to define the payoffs from competition. They are indicators of social or economic impacts on the payoffs. The larger these values, the more benefits for the corresponding language groups. The parameters $\bar{s}_A, \bar{s}_B$ are used to define the payoffs from intervention. They are rates for language reversing due to interventions. The larger these values, the faster the reversing rates.

## 3. Dynamics Analysis

Several theoretical results can immediately be established following the model given in the previous section: (i) Languages $A$ and $B$ can co-exist when the payoffs for speaking $A$ and $B$ are balanced. (ii) The co-existence is evolutionarily stable under certain conditions. (iii) Interventional conditions can be found to maintain an evolutionarily stable multilingual population. The following are more detailed analysis.

Without loss of generality, let $\lambda = 0.5$ and $c = \bar{c} = 1$. Consider a $(y_A^*, y_B^*)$-population, $y_A^*, y_B^* \neq 0$, i.e., languages $A$ and $B$ co-exist in the population. Then, it is easy to see that *a necessary and sufficient condition* for $(y_A^*, y_B^*)$ to be an equilibrium solution to the equations in (3) or in other words, an equilibrium strategy for the game in (6) is $\tilde{P}_A(y_A^*) = \tilde{P}_B(y_B^*)$, i.e.,

$$(y_A^*)^{\alpha-1} s_A + (y_A^*)^{\bar{\alpha}-1} \bar{s}_A = (y_B^*)^{\alpha-1} s_B + (y_B^*)^{\bar{\alpha}-1} \bar{s}_B. \tag{10}$$

For a specific population, for example, for $\alpha = 3/2$ and $\bar{\alpha} = 0$, it can be simplified to

$$(y_A^*)^{1/2} s_A + (y_A^*)^{-1} \bar{s}_A = (y_B^*)^{1/2} s_B + (y_B^*)^{-1} \bar{s}_B. \tag{11}$$

In addition, *a sufficient condition* for $(y_A^*, y_B^*)$ to be evolutionarily stable is

$$(1 - \bar{\alpha})[(y_A^*)^{\bar{\alpha}-2} \bar{s}_A + (y_B^*)^{\bar{\alpha}-2} \bar{s}_B] > (\alpha - 1)[(y_A^*)^{\alpha-2} s_A + (y_B^*)^{\alpha-2} s_B]. \tag{12}$$

For a specific population, for example, for $\alpha = 3/2$ and $\bar{\alpha} = 0$, the condition can be simplified to:

$$(y_A^*)^{-2} \bar{s}_A + (y_B^*)^{-2} \bar{s}_B > [(y_A^*)^{-1/2} s_A + (y_B^*)^{-1/2} s_B]/2. \tag{13}$$

Assume that $(y_A^*, y_B^*)$, $y_A^*, y_B^* \neq 0$, is an equilibrium solution to the system of equations in (3) satisfying the equilibrium condition in (10). Then, several interventional conditions can be obtained to make the solution to be evolutionarily stable when the interventional parameters $\bar{s}_A$ and $\bar{s}_B$ fall in certain ranges:

**Condition 1**: If $1 - \bar{\alpha} \geq \alpha - 1$, i.e., $\alpha + \bar{\alpha} \leq 2$, the stability condition in (12) can be satisfied easily when the reversing rates $\bar{s}_A$ and $\bar{s}_B$ are in certain ranges: Since $\bar{\alpha} - 2 < \alpha - 2$, $(y_A^*)^{\bar{\alpha}-2} > (y_A^*)^{\alpha-2}$ and $(y_B^*)^{\bar{\alpha}-2} > (y_B^*)^{\alpha-2}$, and therefore, the stability condition in (12) is satisfied if $\bar{s}_A$ and $\bar{s}_B$ are sufficiently large, say $\bar{s}_A = ts_A$ and $\bar{s}_B = ts_B$, where $1 \leq t \leq \min\{1/s_A, 1/s_B\}$. With such a choice of $\bar{s}_A$ and $\bar{s}_B$, one can prove that $(y_A^*, y_B^*)$ is also unique (see an example in Figure 1 (a)).

**Condition 2**: If in particular, $1 - \bar{\alpha} = \alpha - 1$, i.e., $\alpha + \bar{\alpha} = 2$, and $\bar{s}_A = ts_B$ and $\bar{s}_B = ts_A$ for any $1 \leq t \leq \min\{1/s_A, 1/s_B\}$, then it is easy to verify that $P_A(y_A^*) = P_B(y_B^*)$ and $\bar{P}_A(y_A^*) = \bar{P}_B(y_B^*)$ if and only if $\tilde{P}_A(y_A^*) = \tilde{P}_B(y_B^*)$, which implies that $(y_A^*, y_B^*)$ is an equilibrium solution to the system of equations in (3) for competition-only and intervention-only and both combined. For the combined one, one can prove that $(y_A^*, y_B^*)$ is also evolutionarily stable.

**Condition 3**: In general, given a desired solution $(y_A^*, y_B^*)$, $y_A^*, y_B^* \neq 0$, it is possible to make it to be an equilibrium solution if the reversing rates $\bar{s}_A$ and $\bar{s}_B$ satisfy the following conditions: For $1 - \bar{\alpha} \geq \alpha - 1$,

$$\bar{s}_A = (y_A^*)^{1-\bar{\alpha}}(y_B^*)^{\alpha-1}s_B, \;\; \bar{s}_B = (y_B^*)^{1-\bar{\alpha}}(y_A^*)^{\alpha-1}s_A. \tag{14}$$

Then, $\bar{P}_A(y_A^*) = P_B(y_B^*)$ and $\bar{P}_B(y_B^*) = P_A(y_A^*)$. It follows that $\tilde{P}_A(y_A^*) = \tilde{P}_B(y_B^*)$, and $(y_A^*, y_B^*)$ becomes an equilibrium solution. In addition, let $y_A^\circ = 1/(1 + (s_A/s_B)^{1/(\alpha-1)})$, and assume that $(y_A^*, y_B^*)$ is selected such that $y_A^* \geq \max\{y_A^\circ, y_B^*\}$ or $y_A^* \leq \min\{y_A^\circ, y_B^*\}$. Then, the condition in (12) is satisfied at $(y_A^*, y_B^*)$, and the solution is also evolutionarily stable (see an example in Figure 1 (b)).

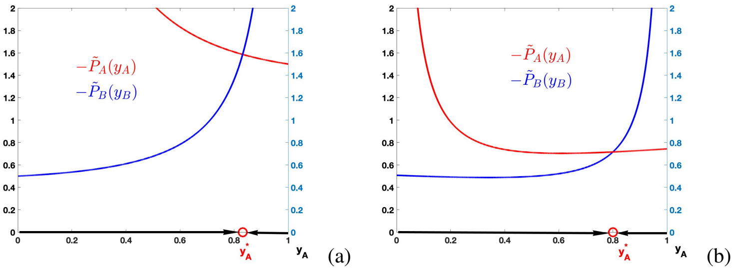


Figure 1. Dynamic Behaviors with Competition and Intervention. Payoff functions $\tilde{P}_A$ and $\tilde{P}_B$ are plotted against $y_A$ and the changing directions of $y_A$ are pointed with arrows. In (a), $\alpha = 3/2$ and $\bar{\alpha} = 0$, $\bar{s}_A = s_A = 0.75$ and $\bar{s}_B = s_B = 0.25$, and $y_A^* = 0.8315$ and $y_B^* = 0.1685$. In (b), $\alpha = 3/2$ and $\bar{\alpha} = 0$, $s_A = 0.6$ and $s_B = 0.4$, $y_A^* = 0.8$ and $y_B^* = 0.2$, and $\bar{s}_A = (y_A^*)(y_B^*)^{1/2}s_B = 0.1431$ and $\bar{s}_B = (y_B^*)(y_A^*)^{1/2}s_A = 0.1073$.

## 4. Computer Simulation

A dynamic simulation is carried out to track the changes of the bilingual level of a population across time and space. A 2D torus-shaped lattice of $n \times n$ cells is constructed first, with each cell assumed to be occupied by an individual speaker. An individual can then be selected repeatedly from the lattice, and a game is played for the individual against the population of the lattice: Let $(x_A, x_B)$ be the current strategy for the individual, and $(y_A, y_B)$ the strategy for the population. Let $p_A = \tilde{P}_A(y_A)$ and $p_B = \tilde{P}_B(y_B)$ be the payoffs for $A$ and $B$ speakers, respectively. Then, the payoff for the individual, $\pi = x_A\tilde{P}_A(y_A) + x_B\tilde{P}_B(y_B)$, is computed. If $p_A > \pi$, $x_A$ is increased by setting $x_A = y_A$ if $x_A < y_A$. On the other hand, if $p_A < \pi$, $x_A$ is reduced by setting $x_A = y_A$ if $x_A > y_A$.

Initially, each individual is assigned with a random strategy. The game is played $n^2$ times for the population to complete a generation. The game is repeated for 100 generations to make sure the population reaches its equilibrium. In general, the game can be played in a neighborhood of each selected individual. Let the neighborhood be an $m \times m$ sub-lattice, with the selected individual located at the center. Then, the game can be carried out for each selected individual only against the population in its neighborhood of this size, with the population strategy $(y_A, y_B)$ computed from the population in the neighborhood. Such a game may in fact be more realistic, as people usually interact only with a small group of others around them.

Figure 2 demonstrates a typical set of results obtained from the simulation for a given population. The population is distributed on a $75 \times 75$ lattice, with $\alpha = 2$ and $\bar{\alpha} = 0$, and $\bar{s}_A = s_B = 0.75$ and $\bar{s}_B = s_A = 0.25$. The simulation is done three times with the neighborhood size equal to $75 \times 75$, $25 \times 25$, and $5 \times 5$, respectively. The final distribution of the individual frequency $x_A^*$ in the population is displayed for each simulation in the corresponding order.

From left to right, the first graph in Figure 2 shows the result from the simulation with the neighborhood size equal to $75 \times 75$, when each individual interacts with all others in the whole population. The equilibrium frequency $y_A^*$ of the population in this case is approximately equal to $0.75$, which agrees with the direct prediction from the model described in previous sections. In addition, the distribution of the individual frequency $x_A^*$ in the population is very homogeneous, with $x_A^* \approx 0.75$ across the board, suggesting that language $A$ and $B$ co-exist in the population in an evenly distributed bilingual form. The second graph shows the result from the simulation with the neighborhood size equal to $25 \times 25$, when the interactions among individual speakers are restricted. The equilibrium frequency $y_A^*$ of the population remains about the same, approximately equal to $0.75$. However, the individual frequency $x_A^*$ becomes less constant. Some regions have higher individual frequencies than others, and local groups are formed with varying individual frequencies, as shown in the graph. The third graph shows the result from the simulation with the neighborhood size further reduced to $5 \times 5$. While the population frequency $y_A^*$ is not significantly changed, the individual frequency $x_A^*$ shows even bigger variations, with even smaller local spots formed with higher or lower individual frequencies than average. The dynamic behaviors shown from these simulations agree with the general experience in language development: Indeed, when communications are restricted to local groups, language variations often remain.

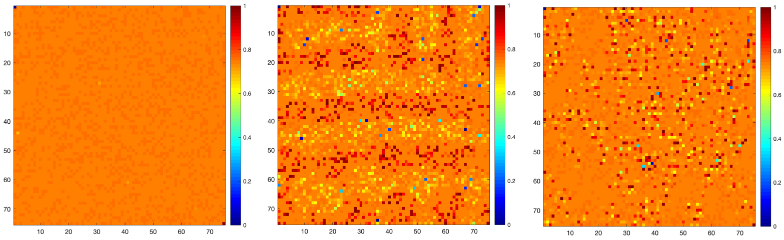


Figure 2. Dynamic Simulation Results. The distributions of color-coded $A$ speaking frequencies in the 2D lattice at equilibrium are displayed in graphs from left to right, with corresponding neighborhood sizes equal to $75 \times 75$, $25 \times 25$, and $5 \times 5$. Each of the graphs is a $75 \times 75$ 2D lattice. The $x$-axis and $y$-axis of the graph represent the 75 units of the lattice in the horizontal and vertical directions. The population is assumed to have $\alpha = 2$ and $\bar{\alpha} = 0$, $\bar{s}_A = s_B = 0.75$ and $\bar{s}_B = s_A = 0.25$.

## 5. Conclusion

An evolutionary dynamics model for multilingual competition with societal intervention is proposed and analyzed. The model consists of two separate parts corresponding to two evolutionary games, one for the evolution of multilingualism with “natural” competition, and the other for the evolution with “artificial” intervention. Both games may have a multilingual co-existing equilibrium state, but the one for the competition-only game is evolutionarily unstable, which leads to the conclusion that multiple languages cannot co-exist, and one of them will eventually dominate while all others become extinct, as stated in many previous studies.

However, multiple languages do exist in many language communities, often in multilingual forms. By combining competition with intervention, the proposed model provides a more general theoretical framework for the study of language competition than those previously investigated. The model shows how multiple languages may co-evolve when appropriate interventions are introduced, and why they may co-exist in stable equilibrium states, at least in theory. The computer simulation on the dynamic behaviors of bilingual populations further validates the model, and also demonstrates how local bilingual groups may be formed when the interactions among the speakers are restricted.

Some experimental work needs to be done to connect the theory to the reality: The parameters in the model need to be refined, denoted, and determined with real-world language data, while their values may vary with varying language populations. For simplicity, the model is defined and discussed only for bilingual populations, but it can in fact be extended to populations with more than two languages, although the analysis may be more mathematically involved (Wu, 2020).

## References

Abrams, D., & Strogatz, S. (2003). Modeling the dynamics of language death. *Nature*, *424*, 900.

Batalova, J., & Zong, J. (2016). Language diversity and English proficiency in the United States. *Migration Information Source*.

Burger, R. (2000). *The mathematical theory of selection, recombination, and mutation.* New York, NY: John Wiley & Sons.

China, P. R. (2017). *China National Language Development Commission: Zhongguo yuyan wenzi shiye fazhan baogao.* Beijing, China: The Commercial Press.

Fujie, R., Aihara, K., & Masuda, N. (2013). A model of competition among more than two languages. *Journal of Statistical Physics*, *151*, 289–303.

Gabszewicz, J., Ginsburgh, V., & Weber, S. (2011). Bilingualism and communicative benefits. *Annals of Economics and Statistics*, *101*, 271–286.

Grosjean, F. B. (2012). *Bilingual: Life and realty.* Cambridge, MA: Harvard University Press.

Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics.* Cambridge, UK: Cambridge University Press.

Mira, J., & Paredes, A. (2005). Interlinguistic similarity and language death dynamics. *European Physics Letter*, *69*, 1031–1034.

Patriaca, M., & Heinsalu, E. (2009). Influence of geography on language competition. *Physics A.*, *388*, 174–186.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707–784.

Shin, H., & Kominski, R. (2010). *Language use in the United States: 2007 – American Community Survey Reports.* Washington D.C.: ACS-12, US Census Bureau.

Vazquez, F., Castello, X., & Miguel, M. S. (2010). Agent based models of language competition: Macroscopic descriptions and order-disorder transitions. *Journal of Statistical Mechanics*, P04007.

Weibull, J. W. (1995). *Evolutionary game theory.* Cambridge, MA: The MIT Press.

Wu, Z. (2020). Why multilingual, and how to keep it – an evolutionary dynamics perspective. *PLoS ONE*, *15*, e0241980.

# ROLE OF MIDBRAIN DOPAMINERGIC SYSTEM IN SOCIAL ENHANCEMENT OF VOCAL LEARNING IN SONGBIRD

SHIN YANAGIHARA*[1], MAKI IKEBUCHI[2], CHIHIRO MORI[3], RYOSUKE O. TACHIBANA[4], KAZUO OKANOYA[2,4]

*Corresponding Author: yanagihara.shin.al@teikyo-u.ac.jp

[1]Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan

[2] RIKEN Center for Brain Science, Saitama, Japan

[3] Department of Molecular Biology, Faculty of Pharmaceutical Sciences,
Teikyo University, Tokyo, Japan

[4] Graduate School of Arts and Sciences, The University of Tokyo, Tokyo,
Japan

There are many similarities between avian vocal learning and human language learning, and in particular, social interaction is known to have a significant impact on vocal learning in both. For example, social interaction is known to facilitate language learning in humans (Kuhl, 2010). Similarly, social interactions between tutor and tutee are also known to enhance vocal learning in songbirds (Chen, Matheson, & Sakata, 2016; Eales, 1989). Juvenile birds learn to sing the same song by listening to the tutor song. In order to learn songs accurately, juvenile birds need to hear the songs directly from their tutor. On the other hand, they cannot learn to imitate songs well by passively listening to songs presented from a speaker. These behavioral studies highlight the importance of social interactions in vocal learning, but the neural mechanisms underlying the social enhancement of vocal learning remain unclear. In this study, we aim to understand the evolutionary origin of language learning by clarifying the neural basis of social enhancement of vocal learning in songbirds. We tested the hypothesis that direct listening to songs from the tutor bird enhances midbrain dopaminergic activity as a social reward for juvenile birds, leading to accurate song memory formation. We measured neural activity in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) of freely behaving juvenile zebra finches. We examined whether auditory responses to the tutor songs changed depending on the presence or absence of a live tutor bird. As a result, we found that a group of neurons in the VTA/SNc showed auditory responses to the tutor song presented from the speaker, and that these responses were markedly enhanced in the presence of the tutor. Moreover, similar enhanced auditory responses were observed when a juvenile listened to a song from a live tutor. These results suggest that midbrain dopaminergic system is involved in the process of juvenile birds learn songs from their tutor.

Next, we tried to find out where in the brain dopamine acts to enhance vocal learning. Previously, it has been suggested that memory of songs learned from a tutor bird is formed in the higher auditory cortex (NCM) (Gobes & Bolhuis, 2007; London & Clayton, 2008; Yanagihara & Yazaki-Sugiyama, 2016). In addition, dopamine receptors are abundant in the higher auditory cortex (Kubikova, Wada, & Jarvis, 2010). Thus, dopamine released in the brain when a juvenile bird hears the song from a tutor may enhance auditory responses in the higher auditory cortex. To test this, we examined whether dopamine modulates auditory responses in the higher auditory cortex. While measuring neural activity in the higher auditory cortex, we found that local administration of dopamine near the recording site markedly enhanced auditory responses to the songs presented from a speaker. These results suggest that dopamine is released into the higher auditory cortex by directly listening to the tutor song, which enhances the auditory responses in the higher auditory cortical neurons and leads to successful song memory formation. In conclusion, we show that dopamine is a key molecule for social enhancement of vocal learning in songbirds. This study provides a new research direction to explore the origin of language through social interaction and dopamine.

## Acknowledgements

This work is supported by MEXT/JSPS KAKENHI 17H06380 (#4903), 17H01015, 20H00105 to K.O. and 17K07066, 20K06907 to S.Y., and ACRO Incubation Grants of Teikyo University.

## References

Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proc Natl Acad Sci U S A, 113*(24), 6641-6646. doi:10.1073/pnas.1522306113

Eales, L. A. (1989). The Influences of Visual and Vocal Interaction on Song Learning in Zebra Finches. *Animal Behaviour, 37*, 507-508. doi:Doi 10.1016/0003-3472(89)90097-3

Gobes, S. M., & Bolhuis, J. J. (2007). Birdsong memory: a neural dissociation between song recognition and production. *Curr Biol, 17*(9), 789-793. doi:10.1016/j.cub.2007.03.059

Kubikova, L., Wada, K., & Jarvis, E. D. (2010). Dopamine receptors in a songbird brain. *J Comp Neurol, 518*(6), 741-769. doi:10.1002/cne.22255

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron, 67*(5), 713-727. doi:10.1016/j.neuron.2010.08.038

London, S. E., & Clayton, D. F. (2008). Functional identification of sensory mechanisms required for developmental song learning. *Nat Neurosci, 11*(5), 579-586. doi:10.1038/nn.2103

Yanagihara, S., & Yazaki-Sugiyama, Y. (2016). Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. *Nat Commun, 7*, 11946. doi:10.1038/ncomms11946

# DISCOVERING HIERARCHY IN AMBIGUOUS COMMUNICATIONS DURING AN EXPERIMENTAL SEMIOTICS TASK

Soma Yano*[1], Naoki Inoue[1], and Junya Morita[1]

*Corresponding Author: yano.soma.18@shizuoka.ac.jp
[1]Faculty of Informatics, Shizuoka University, Hamamatsu, Japan

How is it possible to detect the hierarchical communication that occurs in experimental semiotics? To answer this question, we collected data in an online experiment and developed a method of analyzing lexical and syntactic ambiguities. The results of this analysis indicated the syntactic structure in a case that shows a large gap of ambiguity between profiles that feature different chunking methods. From this, we conclude that our analytical method is useful for detecting hierarchical communication.

## 1. Introduction

Human beings can utilize language to share their intentions, though it is not known how language evolved or how it came to represent the complexity of our thinking. The interdisciplinary field *Evolingusistics: Integrative Studies of Language Evolution for Co-creative Communication* investigates the mechanisms of language evolution using hierarchy and intention sharing as the two main conceptual grounds (Okanoya, 2022). These two concepts capture varying functions of human language. According to Hashimoto (2020), the former concept, which corresponds to linguistic syntax, enables the representation of complex knowledge, whereas the latter indicates an aspect of the semantics and pragmatics of language, leading to knowledge sharing in the language community.

The core question in the field at present regards the integration of the two concepts, in the form of the following question: “How did the hierarchical representation of our thought emerge in relation to the emergence of social collaboration in the history of humanity?” We consider that one possible link can be found in linguistic ambiguity. According to previous studies (Hashimoto, 2020), compound words such as ‘unthinkable’ have led to different interpretations, depending on an applied hierarchical structure such as [[un, think], able] or [un, [think, able]]. Linguistic ambiguity has also been investigated in the semantic and pragmatic domains. Pinker, Nowak, and Lee (2008) find that people may intentionally use ambiguous expressions (e.g., indirect speech) to maintain social relations.

Thus, linguistic ambiguity is involved in two main features of human language. Traditionally, ambiguity that relates to intention sharing and hierarchy has been understood to be lexical and syntactic ambiguities (MacDonald, Pearlmutter, & Seidenberg, 1994). Lexical ambiguity is a reflection of the fact that a single word may have multiple meanings, and syntactic ambiguity relates to the fact that of the possibility of multiple hierarchy.

In this study, we experimentally examine these ambiguities by means of an approach called experimental semiotics, as proposed by Galantucci (2009). In a typical task relating to experimental semiotics, the participants play a collaborative game by exchanging symbols the meanings of which have not been established in advance. This approach allows us to observe language formation in a laboratory setting where the usual means of communication have been deliberately restricted.

As an initial step in exploring the role of ambiguity in hierarchical intention sharing, this study seeks to address the following research question: "How is it possible to detect the hierarchical communication that occurs in experimental semiotics?" To approach this question, this study has identified a novel method of analysis to quantify lexical and syntactic ambiguities using message logs obtained in an experimental semiotic task.

## 2. Method

### 2.1. *Task*

We adopt the task developed by Inoue and Morita (2021). It employs a cooperation and defection dilemma, similar to the social relations that Pinker et al. (2008) identify as a factor that induces linguistic ambiguity. Considerations of space limit this paper to only give an overview of the original task with a modification. As shown in Figure 1, the game consists of repetitive rounds including three independent phases: (i) allocation of objects (two players and one reward) in a $3 \times 3$ grid space, (ii) exchanging messages between two players (only one message for each participant per round), and (iii) the movement of players. Through iterative rounds, players seek to obtain as many rewards as possible. Once a reward is obtained in a monopolized (only one player reaching it) or shared (two players simultaneously reaching it) manner, the next round begins with the objects randomly assigned to new locations. If the both players fail to reach the reward, the next round maintains the location of the objects from the previous one.

As the task belongs to experimental semiotics, the focus of the analysis falls on the second phase. To create a message, the players combine figures whose usage was not shared in advance. Assigning meanings to each combination of figures, players can collaboratively search for the rewards in the grid space. It is efficient for players to share their visibility to reach the reward as they can only make observations of horizontally or vertically adjacent objects (the other player or the reward) from their location.

The above setting, which uses novel communication media, is common in experimental semiotics (Galantucci, 2009). Inoue and Morita (2021) characteris-

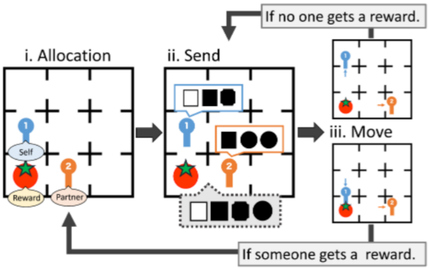


Figure 1. Flow of The Game

tically use discrete figures as message components to allow a simple count-based analysis. In this study, we modified the number of message components, allowing players to horizontally sequence three figures chosen from among four options. In the original task presented by Inoue and Morita (2021), players combine two figures from among ten options. Our modification was made to allow players to use a hierarchical structure when they create messages.

## 2.2. *Participants*

In all, 76 participants (43 men and 33 women) were recruited through a Japanese crowdsourcing site, lancers.jp. They were paid 800 Japanese yen (JPY) for their participation. They also could obtain extra payment depending on the number of rewards obtained in the task.

## 2.3. *Procedure*

The experiment was executed in an online environment developed by the authors. Participants who simultaneously accessed the crowd-sourcing site were paired and executed a collaboration and a dilemma game, in succession. Each game lasted for 20 mins. In the collaboration game, the reward was only obtained when both participants simultaneously reached the reward, whereas in the dilemma game, the participant could obtain the reward in a monopolized or a shared manner. The two manners were weighted in differently in the two conditions: 3:2 and 2:3.[1] After finishing the dilemma game, the participants completed a post-questionnaire (free description), including a prompt to “give your assigned meanings to the figures.”

---

[1] Each number separated by colon indicates points obtained by the share and monopolize manner. These points ware used to calculate extra payment for the participants (obtained points × 3 JPY).

Table 1. Dimensions of connotations.

| Viewpoint | Object | Tense | Intention of the message | Label of axis |
|---|---|---|---|---|
| Absolute | Self | Present | I'm in the middle left now. | AbsSelfPre |
| | | Post | I will go to the lower left next. | AbsSelfPost |
| | Partner | Present | The partner is in the middle right now. | AbsPartnerPre |
| | | Post | I want my partner to come to the middle. | AbsPartnerPost |
| | Reward | - | The reward is in the bottom left. | AbsReward |
| Relative | Self | Post | I will move down next. | RelSelfPost |
| | Partner | Present | My Partner is not adjacent to me. | RelPartnerPre |
| | | Post | I want my partner to come to my left. | RelPartnerPost |
| | Reward | - | The reward is below me. | RelReward |

## 3. Analysis

This study analyzed messages sent during the dilemma game with no distinction between the two experimental conditions. The analysis was conducted to discover hierarchical structures by quantifying lexical and syntactic ambiguities in messages.

### 3.1. *Lexical Ambiguity*

To grasp lexical ambiguity in this task, we made two assumptions. First, we assumed each combination of three figures was a chunk referring to a single location in grid space. Second, we assumed that there were connotations that specified the meanings attached to each denotation (location). Each row of Table 1 indicated an assumed connotation characterized by the three dimensions "viewpoint," "object," and "tense."

The above assumptions indicate that lexical ambiguity can be further classified into two types, corresponding to denotation and connotation. The ambiguity of denotation indicates a weak one-to-one relationship between the locations and combination patterns of the figure. On the other hand, the ambiguity of the connotation represents the degree that one's message presents multiple connotations.

To represent both types of lexical ambiguity, we created profiles of each participant using message logs. This profile consists of nine axes that corresponds to the rows in Table 1. Figure 2 shows how ambiguity of each axis was calculated. The steps taking in this calculation were as follows.

1. *Chose a connotation*: First, one axis from the nine axes in Table 1 is selected.
2. *Count rounds along with possible denotations*: This step counts the rounds that appear in the log data according to the object locations. If the viewpoint is absolute, nine states are defined [(0,0), (0,1), (0,2),... (1,0),... (2,2), where each represents x-y coordinates]. Otherwise, six states are assumed [upper, lower, left, right, same, invisible].

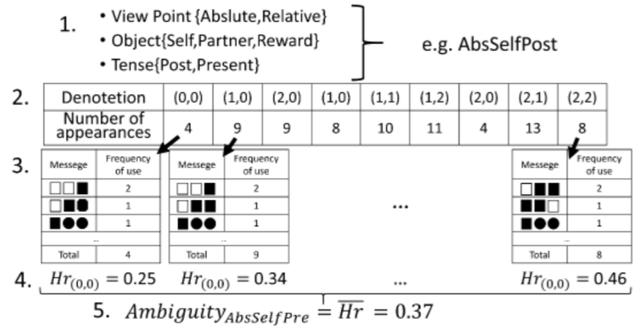


Figure 2. Diagrammatic Illustration of Calculation of Ambiguity

3. *Count figure chunks for each denotation*: This step counts the number of appearances in the log for each combination of figures.

4. *Calculate normalized information entropy for each denotation*: For each denotation (location), this step calculates the ambiguity as normalized information entropy $Hr = \frac{n\log_2 n - \sum_i^k n_i \log_2 n_i}{n} \times \frac{1}{log_2 k}$, where $n$, $n_i$, and $k$ represent the number of appearances of each state, the number using the $chunk_i$, and the number of chunk patterns ($4^3 = 64$), respectively. This index is also normalized through division by $log_2 k$, which is the maximum value of this index.

5. *Aggregating the information entropy across denotations*: In the final step, information entropy is averaged by aggregating the denotation in the axis.

The above ambiguity score represents the ways in which the participant inconsistently uses the chunks of figures under an assumed connotation. We computed this score for each connotation and summarized the results in radar charts.

### 3.2. *Syntactic Ambiguity*

The analysis so far has treated three figures as one chunk. To investigate the structure inside the chunk, we calculated the truncated ambiguity by dividing three figures into two sub-parts: the front two and back two. In the case of there is no difference between the ambiguities calculated from the different sub-parts, we can assume that there is no hierarchical structure. On the contrary, if there is a difference between the two, we consider the possibility of utilizing a hierarchical structure inside the three figures provided by the participant.

## 4. Results and Discussion

To address the research question in the first section, we report representative cases that do not show statistical results.

### 4.1. *Lexical Ambiguity*

Table 2 presents five cases indicating how the profile of lexical ambiguities (the second column) presents the characteristics of the communication system in the task. These cases are chosen based on two indices, and the size and distortion are shown in the third and the fourth columns (the numbers in the parentheses indicate the rank among 76 participants). The size is the average ambiguity across nine axes ($Mean = .271$, $SD = .067$). The distortion marks the difference between the value of the minimum ambiguity and the average ambiguity ($Mean = .081$, $SD = .017$). These cases in the table are chosen to show high and low values of the indices according to the participants who provided interpretable free descriptions in the post-questionnaire (the fifth column).

As can be seen in the upper two cases (ID 1, 2), a large size profile indicates cases where participants conveyed several connotations. On the other hand, smaller profiles indicate a simple message without much consideration (ID 3) or sending mostly the same combinations of the figures (ID 4). The final case (ID 5) shows a highly distorted profile. In this case, ambiguity in the axis labeled AbsSelfPost indicates low values consistent with the free description that can be interpreted as mentioning moving locations.

### 4.2. *Syntactic Ambiguity*

Table 3 shows cases obtained by analyzing syntactic ambiguity. The three profiles, distinguished by the chunking method (three figures as a chunk, with the front two and back two figures as chunks) are overlain in the figure. To find representative cases showing syntactic ambiguity, we calculated the root-mean square difference between the front and the back two profiles (RMSDp) ($Mean = .036$, $SD = .002$). ID 6 is a typical case in this index, while ID 7 shows the third largest value in this index. Such differences in RMSDp seem consistent with the free descriptions. The description by ID 7 can be read as distinguishing the composition of the first two figures from the last to convey complex intentions.

## 5. Summary

This study addressed the question: “How is it possible to detect hierarchical communication occurring in experimental semiotics?” To answer this, we developed a method of analyzing lexical and syntactic ambiguities. This analysis successfully identified syntactic structures in cases involving large a gap of ambiguity between profiles adopting different chunking methods. We consider that our anal-

Table 2. Examples of profiles and free descriptions in the questionnaire

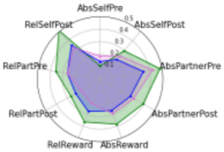
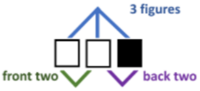
| ID | Profile | Size | Distort | Free descriptions |
|---|---|---|---|---|
| 1 |  | 0.409<br>(2) | 0.064<br>(47) | A quadrangle is 1, a circle is 2, and an octagon is 3, representing the vertical and horizontal axes. Of the three input boxes of the message, if you enter it in the left two, it will show *my* coordinates, if *I* enter it in the center and right, it will show the coordinates of food, and so on. |
| 2 |  | 0.379<br>(4) | 0.062<br>(49) | It showed what was in the room I was in contact with. Each figure *means* a quadrangle: a blank room, a circle: the opponent's piece, an octagon: a reward, a blank: when I was in the four corners, there were two rooms in contact, so I expressed one place as a blank. |
| 3 |  | 0.173<br>(71) | 0.100<br>(17) | At first, I used it *as* the north, south, east, and west in the direction I was going (blank is north, square is east, circle is south, octagon is west), but it *was agreed* from the actions of the two that increasing the number of trials will earn points. From the point on, it didn't make any sense. |
| 4 |  | 0.123<br>(75) | 0.123<br>(9) | There are two types of messages: blank, blank, and blank *shows that I* can't see tomatoes, and blank, circle, and blank *shows that I* can see tomatoes. |
| 5 |  | 0.329<br>(14) | 0.254<br>(1) | From *the dilemma game* , I tried to show where *I am.* The circle is in the first row, the quadrangle in the second row, and the octagon in the third row, with a blank space. Also, if it is all blank, it means that *I am* not moving because it is close. |

*Note:* The free descriptions were transliterated from Japanese using Google Translate. Before the translation, grammatical errors of the original text were corrected and several terms could not been successfully translated were manually corrected (italicized words).

ysis method is useful for detecting compositional communication, a precondition of hierarchical communication (Saldana, Kirby, Truswell, & Smith, 2019).

We consider that the study forms a contribution to the goal of evolinguistics, although the analytical method chosen is limited to a specific type of experimental semiotic task using discrete components of symbols. Using the method presented in this paper, we observe the process of evolving hierarchical communication. In

Table 3. Examples of profiles showing syntactic ambiguity

| ID | Profile | RMSDp | Response to the questionnaire |
|---|---|---|---|
| 6 |  | 0.028<br>(39) | The circle is yourself, the octagon is food, the square is nothing, and the blank is *meaningless*. |
| 7 |  | 0.099<br>(3) | The X-axis 1 2 3 of where you are is circle, octagonal, four strokes, the Y-axis 1 2 3 is square, octagonal, circle, and the *case of staying is blank*. It was difficult to specify whether to proceed if there were two directions. |
| |  | |  |

future work, we will explore the conditions emerging from hierarchical communications using the methods presented in this paper.

## References

Galantucci, B. (2009). Experimental semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science*, *1*(2), 393–410.

Hashimoto, T. (2020). The emergent constructive approach to evolinguistics: Considering hierarchy and intention sharing in linguistic communication. *Journal of Systems Science and Systems Engineering*, *29*(6), 675–696.

Inoue, N., & Morita, J. (2021). A behavioral task for exploring dynamics of communication system in dilemma situations. *Artificial Life and Robotics*, 1–9.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, *101*(4), 676.

Okanoya, K. (2022). *Introduction –evolinguistics: Integrative studies of language evolution for co-creative communication.* (`http://evolinguistics.net/en/`)

Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of sciences*, *105*(3), 833–838.

Saldana, C., Kirby, S., Truswell, R., & Smith, K. (2019). Compositional Hierarchical Structure Evolves through Cultural Transmission: An Experimental Study. *Journal of Language Evolution*, *4*(2), 83-107.

# *HOMO REGULARIS*: LANGUAGE AS A CULTURAL INVENTION OF THE RULE-OBSESSED SPECIES

Masato YOSHIKAWA*[1]

*myshkw@gmail.com
[1]Faculty of Letters, Keio University, Tokyo, Japan

As is often pointed out, human beings are quite distinctive in their astonishing ability to find patterns and rules based on the observation of recurrent phenomena, static sequences and so forth. The ability is so peculiar that they can even *learn* rules that are only assumed to exist but actually do not (e.g., as seen in the result of so-called iterated learning experiments: Kirby & Hurford, 2002). Even more unique is, however, their extraordinary inclination to *obey* the rules they find: not only are they good at detecting recurrent patterns from observation and memorize them, but also enforcing themselves (and others) to follow the patterns, which is characterized as the norm psychology (e.g., Henrich, 2015).

In this sense, human beings can be called *Home regularis*, man of rules: they are unique species in that they can hardly help finding, learning, and obeying rules. This study explores the rule-obsessed nature of human beings and tries to describe the nature of human language as what Homo regularis evolved to have during the course of cultural evolution. Specifically, the "language of Homo regularis" view, the LHR view for short, assumes that 1) language acquisition in the course of ontological development especially concerning grammatical rules is based on the process of *rule generation*, as opposed to rule learning; 2) the basic mechanism of rule generation is *abductive reasoning*, or *abduction*, rather than induction, which is performed by forming possible hypotheses, namely rules, based on input data even if the data are sporadic; 3) the computational engine of rule generation is not a combinatorial operation of minimum units (e.g., words) such as Merge (e.g., Hauser, Chomsky, & Fitch, 2002), but exemplar-based *analogy*.

The best strategy to guess what linguistic rules as social conventions or *norms* are shared within a community would not be to simply generalize relevant input data, because rules are implicit and inputs are concrete while target structure, namely rules, are, or at least can be, abstract. This is a part of the so-called *poverty of stimulus* (POS) argument (e.g., Chomsky, 1980), and as a possible solution to the POS problem, this study assumes that rules are not learned through generalizing input data, but *generated* with the inference from input data. As described above, humans as Homo regularis have a strong tendency to *see* rules in recurrent phenomena even when there are no such rules, which means what they usually do when they think they learn rules would actually be rule generation.

This assumption could reconcile the long-standing nature-nurture dispute of grammatical knowledge/ability. The LHR view presupposes an innate capacity to perform grammatical operation, for example, ordering linguistic units in a semantically consistent way, as evidenced by emergent grammatical patterns found in home signs (e.g., Goldin-Meadow & Feldman, 1977) and the emergence of a new language (e.g., Nicaraguan Sign Language). The view, however, puts more emphasis on the *inheritance* of grammatical conventions shared and maintained in a community throughout the history, so the self-generation of grammatical rules, which seems actually possible, is of second-hand importance. In the course of cultural evolution, humans can generate grammatical rules even when they think they do replicate what the former generation does, as shown in the so-called iterated learning experiments (e.g., Kirby, Cornish, & Smith, 2008). From this it follows that, under the LHR view, language evolution as a cultural evolution is considered to be the repeated process of rule-generation by *pseudo-learners*.

If the language acquisition process is best characterized as rule generation, what mechanism makes the process possible? The answer the LHR view provides to this question is *abductive reasoning*, or *abduction*, which enables us to build a hypothesis about a possible rule or cause generating a result, only based on the observation of the result. Rule-generation could also be equated with the process of *inductive learning* as in, for example, Kirby and Hurford (2002), but inductive process might need a lot more assumptions of learning mechanisms or algorithms than abductive reasoning. Although abduction as the basic mechanism of language acquisition is critically discussed by Chomsky (1968/2006), the objection could be refuted by combining the rule generation process discussed above and the so-called *norm psychology*. Humans are sensitive to others' behaviors and would-be rules behind them but great apes such as chimpanzees do not seem to be (e.g., Haun, Rekers, & Tomasello, 2014). Abductively constructed rules or hypotheses may, therefore, well be effectively rejected by observing others' actions and reactions, resulting in moderately homogeneous behavior and mildly different but largely common sets of generated rules.

If the abduction-based rule generation is the basic mechanism of human language acquisition, how can we construct specific rules based on observed data? Under the LHR view, the computation of rule generation is based on the operation of *exemplar-based analogy* (Cf. Skousen, Lonsdale, & Parkinson, 2002). This marks a stark contrast with the view presupposing as a computational engine of grammar a combinatorial operation such as Merge (e.g., Hauser et al., 2002). The LHR view considers that exemplar-based analogy is more *cost-effective* in the context of computational load than combinatorial operation, in that it does not need full segmentation and can avoid the possible risk of combinatorial explosion (see the supplementary material).

## References

Chomsky, N. (1968/2006). *Language and mind* (Third ed.). Cambridge: Cambridge University Press.

Chomsky, N. (1980). *Rules and representations.* New York: Columbia University Press.

Goldin-Meadow, S., & Feldman, H. (1977). The development of language-like communication without a language model. *Science*, *197*(4301), 401–403.

Haun, D. B., Rekers, Y., & Tomasello, M. (2014). Children conform to the behavior of peers; other great apes stick with what they know. *Psychological science*, *25*(12), 2160–2167.

Hauser, M. D., Chomsky, N., & Fitch, T. W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.* Princeton: Princeton University Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–147). London: Springer.

Skousen, R., Lonsdale, D., & Parkinson, D. B. (Eds.). (2002). *Analogical modeling: An exemplar-based approach to language.* Amsterdam/Philadelphia: John Benjamins Publishing.

# YOUR THEORY OF LANGUAGE EVOLUTION ALSO DEPENDS ON YOUR VIEW OF EVOLUTION

THIAGO MACEK GONÇALVES ZAHN*[1]

*Corresponding Author: thimacek@gmail.com
[1]Departamento de Linguística, Universidade de São Paulo, São Paulo, Brazil

One exciting but often puzzling aspect of the field of Language Evolution is the variety of theories and approaches it includes. These not only embrace different, sometimes contradictory assumptions, but also often seek explanations to different questions. An interesting topic is thus how different theories/ approaches are related, i.e. how different assumptions may lead to disparate questions and views, and how contradictory different assumptions made in the field in fact are.

In an important contribution, Jackendoff (2010) argued that "what there is for a theory of language to explain" depends on one's theory of "what language is", discussing how different views on the "innate language capacity" and on how domain-specific it is lead to different theories on how that language capacity might have evolved. As Jackendoff mentions, some defend that little or nothing special (i.e. domain-specific) is needed for the evolution of language, making it essentially a "cultural phenomenon". In the same article, however, he states that "if that ("disparate languages" passed down through cultural transmission) is all there is to language, a theory of the evolution of language has nothing at all to explain". Therefore, although Jackendoff speaks of cultural evolution, he sees a clear contrast between this and language evolution, which is apparently equated with biological evolution and deemed more important.

Others distinguish even more clearly between *evolution* as a specifically biological phenomenon, and *language change*: Berwick & Chomsky (2016, p. 92) explicitly say "Languages change, but they do not evolve. (…) nonbiological evolution (…) is not evolution at all"; and Andersen (2006) has a whole chapter arguing "that there is no chance of explaining language *change* by the mechanisms of *evolutionary* theory". For some (e.g. Berwick & Chomsky), this distinction is taken as given - perhaps reflecting the Chomskyan view that linguistic variation and change are strongly constrained by a shared Universal Grammar. Others, like Andersen (2006) and Itkonen (1999), have discussed at some length what they see as *disanalogies* that would disallow the use of *evolutionary thinking* to explain *cultural* linguistic phenomena.

Although researchers who advocate for a cultural evolutionary approach to language change and *evolution* (i.e. *origin of modern human languages*) have provided some answers to such criticisms (e.g. Dediu et al, 2013, p. 305-307; Steels, 2017), little conversation seems to occur between both sides of the divide. In part, this may reflect different conceptions of *language*, as pointed by Jackendoff (2010), mirroring the Formalist-Functionalist divide common in linguistics. However, it seems clear that another aspect underlying the divide are different understandings of *evolution* – including *what counts as evolution*, *what evolution can/ should explain* and related questions. This suggests that some of the ongoing debates in the field are *inherently theoretical*, and thus cannot be solved solely empirically, since new evidence may be interpreted differently and/ or given different weights depending on one's assumptions.

Discussion on “what is evolution” and on whether it is possible (or productive) to expand evolutionary thinking beyond biology has a prolific history in philosophy (e.g. Lewontin, 1970; Hull, 1988; Godfrey-Smith, 2007, 2009, 2012; Baraghith & Feldbacher-Escamilla, 2021; also Price, 1995[1971]; Frank, 2012; Luque, 2017). Assuming an ontology based on a ‘general selection theory’ (e.g. Croft, 2000; Clark, 2010; Gong, 2012; Steels & Szathmáry, 2018) allows *linguistic* changes to be studied alongside or independently of genetic changes in speakers, and I suggest it may solve some incommensurabilities between approaches. In fact, models based on such assumptions have shown how *cultural language evolution* could have worked alongside biological evolution to shape modern human language(s) (see e.g. Steels, 2010; Kirby, 2013). Meanwhile, taking ‘evolution’ as limited to *genetic* changes raises questions about what is inherently different in systems of *cultural* change, and whether phenomena at that level might not have long-term effects in the evolution of human language(s).

Taking as a “general model” Kirby (2017, p.125)'s idea that language involves the interaction of three dynamical systems (*individual learning, cultural evolution and biological evolution*), important differences between theories may be recognized by considering which of these systems in fact involve *evolution*, which are deemed more important, and how they influence each other in each approach.

It is perhaps not surprising that one's theory of *Language Evolution* depends on one's theories of *language* and *evolution*; thus, clearing assumptions in both regards may go a great way in building more constructive exchanges in the field.

## Acknowledgments

The author would like to thank João Figueiredo Nobre Cortese, Thomas Finbow, Paulo Chagas de Souza, Joana Bortolini Franco, Wellington Santos da Silva, Mariana Payno and Vitória Lima for constructive discussions about the nature of language and language evolution, and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for financial support which was essential to allow the development of this research.

## References

Andersen, H. (2006) Synchrony, Diachrony, and Evolution. In: O.N. Thomsen (Ed.) *Competing models of linguistic change – Evolution and Beyond* (pp. 59-90). Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Baraghith, K, & Feldbacher-Escamilla, C. (2021). The Many Faces of Generalizing the Theory of Evolution. *American Philosophical Quarterly 58(1)*: 35-50.

Berwick, R., & Chomsky, N. (2016). Why only Us: Language and evolution. Cambridge, MA: MIT Press.

Clark, B. (2010). Evolutionary Frameworks for Language Change: The Price Equation Approach. *Language and Linguistics Compass 4(6)*: 363-376.

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach*. London: Longman.

Dediu, D., Cysouw, M., Levinson, S.C., Baronchelli, A., Christiansen, M.H., Croft, W., Evans, N., Garrod, S., Gray, R.D., Kandler, A., Lieven, E. (2013) Cultural Evolution of Language. In: P.J. Richerson and M.H. Christiansen (Eds.) *Cultural Evolution: Society, Technology, Language, and Religion*. (pp. 303-332). Cambridge, MA: MIT Press.

Frank, S.A. (2012). Natural selection. IV. The Price equation. *Journal of Evolutionary Biology* 25: 1002-1019.

Godfrey-Smith, P. (2007). Conditions for Evolution by Natural Selection. *Journal of Philosophy* 104: 478-516.

Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. New York, NY: Oxford UP.

Godfrey-Smith, P. (2012). Darwinism and cultural change. *Philosophical Transactions of the Royal Society B 367*: 2160-2170.

Gong, T., Shuai, L., Tamariz, M., Jäger, G. (2012). Studying Language Change Using Price Equation and Pólya-urn Dynamics. *PLoS ONE 7(3)*: e33171.

Hull, D.L. (1988). A mechanism and its metaphysics: An evolutionary account of the social and conceptual development of science. *Biology and Philosophy* 3(2): 123-155.

Itkonen, E. (1999). Functionalism yes, biologism no. *Zeitschrift für Sprachwissenschaft, 18(2)*: 219-221.

Jackendoff, R. (2010). Your theory of language evolution depends on your theory of language. In: R.K. Larson, V. Déprez, H. Yamakido (Eds.) *The Evolution of Human Language – Biolinguistic Perspectives* (pp. 63-72). Cambridge: Cambridge University Press.

Kirby, S. (2013). Transitions: The Evolution of Linguistic Replicators. In: Binde, P.-M.; Smith, K. (Eds). *The Language Phenomenon: Human Communication from Milliseconds to Millennia*. (pp. 121-138). Heidelberg: Springer.

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin and Review 24*: 118-137.

Lewontin, R.C. (1970). The units of selection. *Annual Review of Ecology and Systematics 1*: 1-18.

Luque, V.J. (2017). One equation to rule them all: a philosophical analysis of the Price equation. *Biology & Philosophy 32(1)*: 97-125.

Price, G.R. (1995[1971]). The Nature of Selection. *Journal of Theoretical Biology* 175: 389-396.

Steels, L. (2010). Can Evolutionary Linguistics Become a Science? *Journal for Evolutionary Linguistics* 1(1): 1-34.

Steels, L. (2017). Do languages evolve or merely change? *Journal of Neurolinguistics 43(B)*: 199-203.

Steels, L., Szathmáry, E. (2018). The evolutionary dynamics of language. *Biosystems 164*: 128-137.

# IS REDUNDANCY USEFUL IN LANGUAGE? AGENT-RECIPIENT DISAMBIGUATION IN ENGLISH AND DUTCH

EVA ZEHENTNER*[1] and DIRK PIJPOPS[2]

*Corresponding Author: eva.zehentner@es.uzh.ch
[1]English Department, University of Zurich, Zurich, Switzerland
[2]Department of Modern Languages, University of Liège, Liège, Belgium

This paper discusses the competing evolutionary motivations of efficiency versus robustness in language processing and learning (MacWhinney et al. 2014), both from a typological and diachronic perspective. Specifically, we assess the potential benefits or costs of redundancy in morphosyntactic marking of participant roles, comparing and testing two opposing hypotheses:

On the one hand, following the most crucial tenet in usage-based linguistics that language use affects – or even determines – grammar (Bybee 2010), we assume that language is organised in a way that facilitates efficient usage (e.g. Gibson et al. 2019). On this account, redundant marking should be dispreferred. Well-known typological 'trade-off' distributions and diachronic trajectories between word order and morphological case marking seem to support this point (Fedzechkina et al. 2017). Furthermore, prepositional marking is often only applied in contexts where it comes with some added processing benefit (cf. Pijpops et al. 2018 on the impact of complexity on Dutch transitive object marking, or Tal et al. 2020, Levshina 2021 on ambiguity/atypicality in differential object marking).

On the other hand, however, we pursue Van de Velde's (2014) argument that a certain amount of redundancy – or rather, 'degenerate' marking (involving many-to-many relationships) – is in fact beneficial from a usage perspective: redundancy constitutes an indispensable component of any degenerative Complex Adaptive System, and thus also of language (Steels 2000; Beckner et al. 2009). Such redundancy/degeneracy comes with two important advantages, viz. robustness and evolvability: most importantly for the present paper, the former entails that redundant marking offers protection against information loss in the noisy language channel, even though it may be less efficient. Redundancy is furthermore assumed to increase learnability, particularly in more complex situations (e.g. Tal et al. 2021).

Our case study to assess the plausibility of what we call the 'strict-efficiency' versus the 'robustness' account is participant role marking in ditransitive clauses in Present Day Dutch and English, for a comparative perspective, as well as historical English for a diachronic view. More precisely, we investigate the interaction between strategies used to distinguish agents and recipients in transfer-events, e.g. with verbs of giving as in (1) and (2).

(1) They$_{\text{AGENT}}$ give a book to the student$_{\text{RECIPIENT}}$.

(2) Ze $_{\text{AGENT}}$ geven een boek aan de student$_{\text{RECIPIENT}}$.

Since both agents and recipients in ditransitive clauses are prototypically animate (sentient) and volitional (e.g. Newman 1998; Naess 2007; Haspelmath 2015), disambiguating these roles based on semantic-pragmatic information is usually difficult if not impossible. Morpho-syntactic cues are hence indispensable in determining 'who gave what to whom'. Among the strategies language users have at their disposal are (i) constituent order (e.g. SVO in Present Day English), (ii) case marking/ formal differentiation (e.g. subject vs object pronoun forms in PDE), (iii) subject-verb agreement, and (iv) prepositional marking. Employing multiple strategies at the same time constitutes redundant marking; for example, in (1) all four disambiguation strategies are given. Meanwhile in (3), none are used, resulting in an ambiguous sentence.

(3) Mijn baas kan je niet zomaar een uitbrander geven.
'You can't just give my boss a telling-off' or 'My boss can't just give you a telling off.'

In our study, we make use of the *Sonar Corpus of Written Dutch* (Oostdijk et al. 2013), a pre-compiled dataset of ditransitives from the *ICE-GB* (Röthlisberger 2018) and the *Penn Parsed Corpus of Middle English* (PPCME2; Kroch et al. 2000). Instances of ditransitive clauses with *give* are extracted from the corpora, and coded for the strategies instantiated by them. Following the 'strict-efficiency' account, we then expect language users to prefer employing a single strategy for each instance. By contrast, based on the degeneracy/ robustness account, we anticipate sentences that simultaneously instantiate multiple strategies to be most common, and cases where only one strategy is at work to be rare. Our results indicate that even though the precise strategies and their disambiguation power differ between Dutch and English, both languages show substantial redundancy to be the default. Still, redundancy seems to operate within limits, with four-fold strategy use being rare, and two simultaneous strategies being most common. Our diachronic results are in line with this conclusion: We find that English appears to have moved towards more redundant marking over time, but that after a short period of 'exuberant' redundancy, double redundancy is settled on as the norm.

In a final step, we assess the question of whether redundant marking is particularly frequent in complex environments, here measured as sentence length in words (excluding the subject and object arguments of the respective ditransitive patterns). Our findings are again mixed: for Dutch and historical English, complexity emerges as an influential predictor; in Present Day English, however, no significant effect can be observed. We interpret this outcome of our study in light of the differing degrees of variability of strategies in the languages/ stages.

## References

Beckner, C., Blythe, R., Bybee, J., Christiansen, M., Croft, W., Ellis, N., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a Complex Adaptive System: Position paper. *Language Learning, 59(1),* 1-26.

Bybee, J. (2010). *Language, usage and cognition.* Cambridge: Cambridge University Press.

Gibson, E., Futrell, R., Piantadosi, S.P., Dautriche, I., Mahowald, K., Bergen, L. & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 23(5),* 389-407.

Fedzechkina, M., Newport, E., & Jaeger, F. (2017). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science, 41(2),* 416-446.

Haspelmath, M. (2015). Ditransitive constructions. *Annual Review of Linguistics 1,* 19-41.

Kroch, A., Taylor, A., & Santorini, B. (2000-). The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Second edition, release 4.

MacWhinney, B, Malchukov, A., & Moravcsik, E. (eds.). (2014). *Competing motivations in grammar and usage*. Oxford: Oxford University Press.

Naess, A. (2007). *Prototypical transitivity.* Amsterdam: Benjamins.

Newman, J. (1998). Recipients and ‘give’ constructions. In W. van Langendonck and W. Van Belle (Eds.), *The dative, Vol. 2: Theoretical and contrastive studies* (pp. 1-28). Amsterdam: Benjamins.

Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013). The construction of a 500-million-word reference corpus of contemporary written Dutch. In P. Spyns and J. Odijk (Eds.), *Essential speech and language technology for Dutch, Theory and applications of Natural Language Processing* (pp. 219-247). Heidelberg: Springer.

Pijpops, D., Speelman, D., Grondelaers, S., & Van de Velde, F. (2018). Comparing explanations for the Complexity Principle. Evidence from argument realization. *Language and Cognition, 10(3),* 514-543.

Röthlisberger, M. (2018). *The dative dataset of World Englishes*. KU Leuven.

Steels, L. (2000). Language as a Complex Adaptive System. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. Merelo and H.-P. Schwefel (Eds.), *Proceedings of PPSN VI* (pp. 17-26). Berlin: Springer.

Tal, S., E. Grossman, H. Rohde & Arnon, I. (2021). Speakers use more redundant referents with language learners: Evidence for communicatively-efficient referential choice. Preprint. *PsyArXiv*.

Tal, S., Smith, K., Culbertson, J., Grossman, E., & Arnon, I. (2020). The impact of information structure on the emergence of differential object marking: An experimental study. Preprint. *PsyArXiv*.

Van de Velde, F. (2014). Degeneracy: the maintenance of constructional networks. In R. Boogaart, T. Colleman and G. Rutten (Eds.), *Extending the scope of Construction Grammar* (pp. 141-179). Berlin: Mouton de Gruyter.

# CATEGORICAL PERCEPTION AND DISCREETNESS: AN EVOLUTIONARY AND NEUROCOGNITIVE PERSPECTIVE

ELIZABETH QING ZHANG 1*[1], EDWARD RUOYANG SHI 2[2]

*Corresponding Author: zqelizabeth@gmail.com
[1] Center for Language Evolution Studies, University Centre of Excellence IMSErt,
Nicolaus Copernicus University, Torun, Poland
[2] Department of Catalan Philology and General Linguistics, University of
Barcelona, Barcelona, Spain

As one of the design features of human language, discreteness is present at both phonological and syntactic levels. We hypothesize that categorical perception (CP) could have paved the way for discreteness across domains including language. Evidence from comparative and neurobiological perspectives are provided in this paper. Comparative studies on CP suggest that it has a phylogenetic root and is a combination of nature and nurture. Neurobiological studies suggest that the connection between auditory cortex and (pre)motor cortex, and between (pre)motor cortex and the basal ganglia could play an important role in CP and further discreteness. This paper discusses how domain general perceptual ability—categorical perception—could have contributed to discreteness, highlighting the role of sensory-motor system in the evolution of language.

## 1. Introduction

In this paper, we propose that Categorical Perception (CP) could have played a foundational role for discreteness of language, which is a prominent feature of human language (Hockett, 1960). We will approach discreteness from domain general perspective and highlight how it is salient in language. Then by reviewing CP of sounds in nonhuman animals, we argue that CP has its phylogenetic roots in terms of evolution. Following this, we explicate how CP could have been the basis for discreteness with neurological evidence focusing on auditory cortex, (pre)motor cortex and the basal ganglia. The current work discusses the role of perception in language evolution, which provides new avenue to explore the evolution of human language from sensory-motor system.

## 2. Discreteness

Discreteness is an essential concept in several areas, such as quantum physics and chemistry. As Hauser et al. (2002) pointed out, life is constructed on separate and unblendable components (genes). Besides, an example in our daily lives is that time is calculated as discrete in the units of seconds. In terms of human cognition, it is generally established that there exist discrete units of mathematics, such as integers, and discrete units of music, such as melody and rhythm. In the domain of human language, as Hockett (1960) pointed out, discrete and compositional language differs from continuous and holistic nonhuman animal communication signals, which he considered as one of the design features of human language. But what Hockett has emphasized is phonological structure of language, namely syllables are made up of separate vowels and consonants at the phonological level. However, discreteness is also observed at syntactic levels. Through merging lexical items into grammatical phrases or sentences at the syntactic level, lexical elements are discretely dispersed and organized into grammatical phrases or sentences. These discrete atoms at various levels of language combine to form bigger structures based on the rules at that level. For example, you may find 6- and 15-word sentences, but not 7.5-word ones. Just as Searle (2008; p 176) defined, discreteness is "the property through which syntactical elements keep their identity when subjected to diverse syntactical processes".

## 3. Categorical perception and Discreteness: Comparative evidence

Categorical perception (CP) is a psychophysical process in which continuous inputs are perceived discretely across modalities (Harnad, 1987). The interactions between low-level perception and high-level cognition are revealed by this implicit segmentation of continuous physical inputs (Goldstone & Hendrickson, 2010). The mechanism of CP has been investigated utilizing psychophysical approaches such as labeling and discerning continuous stimulus paradigms (Harnad, 1987). The phenomenon of CP was initially identified in sound modality in human speech sounds (Liberman et al., 1957), where listeners were prone to perceive the b-d-g continuum into three distinct groups. Furthermore, research on infants has revealed that they appear to be endowed with the ability to discriminate different sounds in all languages (which is the nature part), but as they have more contact with one (or more for multilinguals) ambient language(s), they tend to group sounds that are not contrastive in their native language(s). For instance, after 8-10 months, Japanese newborns can no longer distinguish between /l/ and /r/, despite the fact that these two sounds are acoustically distant. Statistical learning (Maye et al., 2002) has been proposed to achieve such postnatal shaping CP of phonemes, which appears to be reliant on the requirement that speech sounds are discrete. Statistical learning is the process of learning the statistical distribution of environmental inputs via sheer exposure to the stimuli without instruction (Romberg & Saffran, 2010). The categories boundaries of the stimuli could be changed as a result of statistical learning. The finding of Japanese newborns no longer discriminating between /l/ and /r/ might be due to implicit statistical learning of the non-contrastive distribution of /l/ and /r/ in Japanese. Statistical learning may also be seen in the process of segmenting words from speech streams (Saffran et al., 1996), which underlies CP in the sense that discrete elements are extracted from continuous inputs. It is also worth mentioning that statistical learning is used in a variety of cognitive domains other than language (Bogaerts et al., 2020).

It is also worth emphasizing that CP is not exclusive to speech sounds, but rather a phenomenon observed in a variety of domains and modalities. If our hypothesis that discreteness arises from CP is correct, the discovery of CP in nonhuman animals shows that discreteness of language may have been derived from a preserved trait, namely CP. Comparative studies in nonhuman animals reveal that CP is phylogenetically anchored in early invertebrates and appears to be a combination of nature and nurture (Zhang et al., 2021). The psychophysical paradigm was also employed to investigate CP in nonhuman animals, and the participants were instructed to discriminate and categorize the stimuli, much as it did in human research. We present examples of CP in the sound modality in nonhuman animals.

Crickets have been observed to be able to distinguish between communication calls and predator ultrasounds at a sharp border when it comes to naturally produced noises (Wyttenbach et al., 1996). Female tungara frogs respond to mating sounds in a categorical manner (Baugh, Akre & Ryan, 2008). These findings imply that sound CP is linked to reproduction and survival. In addition, early research looked at how nonhuman animals categorize human speech sounds, which demands a high level of auditory learning capacity. Nonhuman primates were shown to be able to discriminate both consonants and vowels (e,g. Sinnott & Mosteller, 2001). Furthermore, the CP of sounds in birds was studied from the standpoints of both conspecific calls and human speech sounds. It was discovered as early in the 1980s that budgerigars not only had a low discrimination frequency corresponding to their contact calls, but also had a similar range of voice onset time to humans in the perceptual change of bilabial, alveolar, and velar continua (Dooling et al. 1987). The budgerigars used the same cues as humans to distinguish between vowel groups (Dooling & Brown, 1990). This shows that CP of speech sounds is related to auditory perception in general. Moreover, the vocal learning continuum hypothesis, with the most widely distributed species having auditory learning ability and a small number of species having vocal production learning ability, is in some ways paralleled by the large-range distribution of categorical perception in sound modality in diverse animals (Jarvis, 2019).

How might CP of sounds be linked to the discreteness of language? To perceive the world in a discrete way seems to be a prerequisite for discrete production. Then how is the discreteness of sounds related to the discreteness of lexical items or words? In this context, one of the most intriguing and contentious questions is how lexical items or words emerged in evolution. Two scenarios have been considered: One is that words evolved directly from animal communication; and the other one is that words evolved from some other cognitive ability. Some researchers have defended that, if language developed from a propositional and holistic animal communication signal, there must have been a period when humans successfully extracted pieces from certain expressions that could be reused in other expressions (Tallerman, 2007), gradually forming compositionality. Another possible scenario is that words developed through vocal imitation of natural sounds. Word learning, according to Hauser et al. (2002), necessitates the ability of vocal imitation, which has been observed in nonhuman species such as songbirds (p. 1574). Recent research found that generations of vocal imitation of some natural sounds can lead to word-like outputs with category labels, implying that vocal imitation may be a major factor in the emergence of first words (Edmiston et al., 2018). Returning to our hypothesis that CP prepared the basis for discreteness, it has been shown that the discreteness of sounds and the discreteness of words are inextricably intertwined. When it comes to CP of speech sounds, or how to discriminate and categorize speech sounds in a given language, it appears that sequential statistics manifested in words (Transitional Probability reference), rather than acoustic features of the sounds, drive discrimination and categorization of the speech sounds (Saffran et al., 1996). Furthermore, data in infants suggest that perceptual statistical learning also plays a key role in word segmentation from speech streams (Romberg & Saffran, 2010). In this sense, the key of discreteness offered at the phoneme level appears to be influenced by how sound sequences in words are organized (Bidelman & Lee, 2015). This shows that discreteness may have been founded on categorical perception.

## 4. Categorical perception and Discreteness: Neurocognitive considerations

Neurobiological studies on CP and discreteness of language also support our proposal. Using birds as an example, HVCx (HVC is a letter-based term) cells in swamp sparrows have been demonstrated to respond robustly to auditory categorical changes in note duration (Prather et al., 2009). In birds, HVC projects to many brain regions, with HVCx being the projection to AreaX (striatal area x) among others. HVC is a premotor nucleus that has been proposed to be analogous to Broca's area in humans and serve sensory-motor functions (Prather et al., 2017). As a result, HVCx could be analogous to the premotor-striatal connection, which has been linked to beat perception in humans (Grahn & Rowe, 2009). Furthermore, the auditory nuclei field L projecting to CLM (caudolateral mesopallium) and NCM (the caudal part of the medial nidopallium), both projecting to CMM (the caudal part of the medial mesopallium), present an analogous hierarchy to humans in which physical information is processed at the lower level while abstract concepts are encoded at the higher level in European Starlings (Jeanne et al., 2011). It is worth noting that NCM and CMM, which are similar to human auditory cortex, have also been documented for auditory memory (Bolhuis & Gahr, 2006). Moreover, Lampen et al. (2017) discovered that NCM and CMM in zebra finches are more responsive to rhythmic than arrhythmic songs, indicating that NCM and CMM are involved in auditory detection and discrimination.

The auditory cortex is also shown to be involved in CP of speech sounds in primates. Spiking activity from the superior temporal gyrus (STG) in rhesus monkeys were recorded (Tsunada et al., 2011). In humans, however, CP of sounds was mediated by not only the STG and superior temporal sulcus (STS) (Harinen & Rinne, 2013), but also the premotor and primary motor cortex (Chevillet et al., 2013) comparable to those found in the NCM and CMM of birds. Furthermore, Hickok & Peoppel (2004) proposed a dual-stream theory, with a ventral stream dealing with phoneme and lexical recognition and lexical combinations, and a dorsal stream dealing with sensorimotor transformation in speech output. The dorsal stream connects the STG with the premotor cortex, both regions were reported to be where CP takes place. If CP is the basis for discreteness, then such dorsal connection could have played a role in discreteness. The dorsal pathway has been shown to be necessary for vocal imitation, which has been shown to be one of the key factors for word emergence (Edmiston et al., 2018). In this sense, discreteness seems to rely on CP in both phonological level, namely phonemes, and syntactic level, namely words.

Beyond cortex, recent evidence has shown that the basal ganglia also relate to cognitive functions beyond motor domain which includes perceptual categorization (e.g., Ashby et al., 1998; Hochstenbach et al., 1998; Seger, 2008, for reviews). Though basal ganglia-mediating category learning and speech perception and learning have been largely studied independently, as Lim et al. (2014) reviewed, these separate lines of research share commonalities and there is great potential in bridging efforts to understand speech perception and learning with general cognitive neuroscience approaches and neurobiological models of Learning. Due to the fact that speech perception required the integration of acoustic cues across various dimensions, explicit attempts to the properties of speech are difficult. It has been shown that explicit/directed attention to acoustic features are engaged in the process of directed categorization training (Logan et al., 1991). Further, in an fMRI study, Tricomi et al. (2006) showed that category training of non-native speech categories engages the basal ganglia (i.e., the striatum). These results suggest that the basal ganglia learning system are involved in promoting adult speech category learning, which is the nurture part of CP. This is also in parallel with the finding in swamp sparrow mentioned above that Area X is also involved in CP of note duration.

## 5. Conclusion

In this paper, we come up with the hypothesis that categorical perception (CP) could have laid the foundation for discreteness, one of the design features of language. By reviewing comparative studies on CP in nonhuman animals and humans, we found that CP has a phylogenetic root dating back to invertebrates which is closely related to reproduction and survival, and seems to be a combination of innateness and experience. In addition, by reviewing neurobiological studies, we show that tasks of CP activate cortical and subcortical areas including auditory and (pre)motor cortex as well as the basal ganglia, the connection between which could be insightful for locating domain general discreteness. The current work provides additional evidence for the important role of sensory-motor system in language evolution.

## References

Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological review,* 105(3), 442.

Baugh, A. T., Akre, K. L., & Ryan, M. J. (2008). Categorical perception of a natural, multivariate signal: Mating call recognition in túngara frogs. *Proceedings of the National Academy of Sciences of the United States of America,* 105(26), 8985‑8988.

Bogaerts, L., Frost, R., & Christiansen, M. H. (2020). Integrating statistical learning into cognitive science. *Journal of Memory and Language,* 115(August), 104167.

Bolhuis, J. J., & Gahr, M. (2006). Neural mechanisms of birdsong memory. *Nature Reviews Neuroscience,* 7(5), 347‑357.

Bidelman, G. M., & Lee, C. C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *NeuroImage*, 120, 191‑200.

Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Annals of Internal Medicine,* 158(6), 5208‑5215.

Dooling, R. J., & Brown, S. D. (1990). Speech perception by budgerigars (Melopsittacus undulatus): Spoken vowels. *Perception & Psychophysics*, 47(6), 568‑574.

Dooling, R. J., Brown, S. D., Park, T. J., Okanoya, K., & Soli, S. D. (1987). Perceptual organization of acoustic stimuli by budgerigars (Melopsittacus undulatus): I. Pure tones. *Journal of Comparative Psychology*, 101(2), 139‑149.

Edmiston, P., Perlman, M., & Lupyan, G. (2018). Repeated imitation makes human vocalizations more word-like. *Proceedings of the Royal Society B: Biological Sciences*, 285(1874).

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science,* 1(1), 69-78.

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *The Journal of Neuroscience,* 29(23), 7540‑7548.

Harinen, K., & Rinne, T. (2013). Activations of human auditory cortex to phonemic and nonphonemic vowels during discrimination and memory tasks. *NeuroImage,* 77, 279‑287.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In Harnad, S. (ed.) *Categorical Perception: The Groundwork of Cognition.* New York: Cambridge University Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve?. *Science*, 298(5598), 1569-1579.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203(3), 89‑96.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.

Jeanne, J. M., Thompson, J. V., Sharpee, T. O., & Gentner, T. Q. (2011). Emergence of learned categorical representations within an auditory forebrain circuit. *Journal of Neuroscience*, 31(7), 2595‑2606.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science,* 9(2), F13-F21.

Lampen, J., McAuley, J. D., Chang, S.-E., & Wade, J. (2017). Neural activity associated with rhythmicity of song in juvenile male and female zebra finches. *Behavioural Processes*, 163, 45-52.

Liberman, A. M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America,* 29(1), 117-123.

Lim, S. J., Fiez, J. A., & Holt, L. L. (2014). How may the basal ganglia contribute to auditory categorization and speech perception?. *Frontiers in neuroscience, 8*, 230.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America,* 89(2), 874-886.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111.

Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., & Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nature Neuroscience*, 12(2), 221‑228.

Prather, J. F., Okanoya, K., & Bolhuis, J. J. (2017). Brains for birds and babies: Neural parallels between birdsong and speech acquisition. *Neuroscience and Biobehavioral Reviews*, 81, 225‑237.

Hochstenbach, J., van Spaendonck, K. P., Cools, A. R., Horstink, M. W., & Mulder, T. (1998). Cognitive deficits following stroke in the basal ganglia. *Clinical Rehabilitation*, 12(6), 514-520.

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science,* 366(6461), 50‑54.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906‑914.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606‑621.

Searle, J. R. (2008). What is language: Some preliminary remarks. *Explorations in Pragmatics: Linguistic, Cognitive and Intercultural Aspects*, 7‑37.

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience & Biobehavioral Reviews*, 32(2), 265-278.

Sinnott, J. M., & Mosteller, K. W. (2001). A comparative assessment of speech sound discrimination in the Mongolian gerbil. *The Journal of the Acoustical Society of America*, 110(4), 1729‑1732.

Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua*, 117(3), 579‑604.

Tricomi, E., Delgado, M. R., McCandliss, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of cognitive neuroscience,* 18(6), 1029-1043.

Tsunada, J., Lee, J. H., & Cohen, Y. E. (2011). Representation of speech categories in the primate auditory cortex. *Journal of Neurophysiology,* 105(6), 2634‑2646.

Wyttenbach, R. A., May, M. L., & Hoy, R. R. (1996). Categorical perception of sound frequency by crickets. *Science,* 273(5281), 1542‑1544.

Zhang, Q., Lei, L., & Gong, T. (2021). Categorical Perception as a Combination of Nature and Nurture. Proceedings of the Annual Meeting of the Cognitive Science Society, 43.

# CROSS-LINGUISTIC EVIDENCE FOR DEPENDENCE BETWEEN THE PERCEPTUAL DISPERSION AND SALIENCE OF VOWEL SYSTEMS

MENGHAN ZHANG*[1,2,3,4] and TAO GONG[5]

*Corresponding Author: mhzhang@fudan.edu.cn
[1]Institute of Modern Languages and Linguistics, Fudan University, Shanghai, China
[2]Ministry of Education Key Laboratory of Contemporary Anthropology,
Department of Anthropology and Human Genetics, School of Life Sciences,
Fudan University, Shanghai, China
[3]Key Innovation Group of Digital Humanities Resource and Research,
Shanghai Normal University, Shanghai, China
[4]School of Foreign Languages, Zhejiang University of Finance & Economics,
Hangzhou, China
[5]Google, 111 8th Avenue, NY 10011

Investigating typological variations of languages worldwide and generalizing their universality is an important research goal in linguistics. In general, the evolution of spoken language is carried out toward the functional goal of improving communicative effectiveness between speakers and listeners. During this process, several functional pressures from perceptual dispersion (Liljencrants and Lindblom, 1972), perceptual salience (Schwartz et al., 2007), and articulatory costs (Evans & Levinson, 2009; Hammarström, 2016; Chirkova & Gong, 2014) have been proposed as the intrinsic driving forces to lead to the universal patterns of sound systems of human speech, especially vowel systems. Following the principle of clarity (Zuidema & de Boer, 2006), the dispersion theory states that the sound systems of human speech are shaped predominantly by perceptual distinctiveness between sounds (Liljencrants and Lindblom, 1972). Moreover, the dispersion-focalization theory integrates both perceptual dispersion and salience of vowels to investigate the general formation mechanism underlying the global structural variation of vowel systems (Schwartz et al., 2007). Moreover, both perceptual dispersion and salience of vowels can be measured based on the formant frequencies of vowels which are related to the tongue movements and opening/closing of the jaw (Oudeyer, 2006; Bradlow, 1995). In addition, the Quantal theory (Stevens, 1972) and the articulatory complexity theory (Lindblom & Maddieson, 2006; Kingston, 2007) characterize articulatory difficulty in speech production where the auditory pressures of maximizing perceptual distance and minimizing articulatory effort cohesively shape the articulatory subspace of a sound system.

Despite these theories being proposed independently, they imply the association between perceptual dispersion and salience under articulatory effort minimization. In this study, we conduct a cross-linguistic comparison using a worldwide phonetic database (Becker-Kristal, 2010). The database contains formant frequencies of each vowel for 532 individual samples covering 357 languages from 36 language families. We adopt two properties named Effective Dispersion estimate (EDE) and Focalization estimate (FE) to capture the structural crowdedness of a vowel system and the spectral salience of intra-vowel sounds, respectively. EDE is specifically a composite property of the classic Dispersion estimate divided by articulatory space area. All properties are calculated on the psychoacoustic domain.

Using a linear mixed effects model (Quené & Bergh, 2008), we identify that the association between perceptual dispersion and salience follows a power-law-based dependence across vowel systems of worldwide languages. We also demonstrate that such dependence was a language-universal tendency, independent of geographic regions, language families, and linguistic affiliations. Using the phylogenetic methods such as PIC (Felsenstein, 1985; Paradis, 2012) and PGLS (Paradis, 2012), we take Indo-European languages as an example to elaborate that the dependence resulted from correlated evolutions of the two structural properties. The results show the correlated evolution could follow an adaptation process under the stabilizing selection pressures. These results also indicate that the evolution of vowel systems could be a dynamic process of adaptive organization of their structures and optimization capacities. The phylogenetic analysis also shows that the dependence should proceed in a punctuated equilibrium. In other words, the correlated evolution of EDE and FE is not a gradual process. It suggests an evolution scenario: vowel systems tend to slowly accumulate structural changes till reaching a threshold, and then, the whole system is rapidly restructured. To extend the scope of early explanations of language universals, we provide our understanding of these findings from four aspects: language change dynamics, self-organizing criticality of system complexity, and the principle of the least effort in human behavior.

## References

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48(4),* 839-862.

Schwartz, J., Boë, L., & Abry, C. (2007). Linking dispersion-focalization theory and the maximum utilization of the available distinctive features principle in a perception-for-action-control theory. In: M-J., Sole, P. S. Beddor, and M. Phala (Eds.), *Experimental Approaches to Phonology* (pp. 104-124). Oxford: Oxford University Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429-492.

Hammarström, H. (2016). Linguistic diversity and language evolution. *Journal of Language Evolution*, *1,* 19-29.

Chirkova, K., & Gong, T. (2014). Simulating vowel chain shift in Xumi. *Lingua, 152*, 65-80.

Zuidema, W., & de Boer, B. (2006). The evolution of combinatorial phonology. *Journal of Phonetics*, *37*, 125-144.

Oudeyer, P-Y. (2006). *Self-organization in the evolution of speech*. Oxford: Oxford University Press.

Bradlow, R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, *97*, 1916.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In: David, E. E. Jr. and Denes, P. B. (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw-Hill.

Lindblom, B., & Maddieson, I. (2006). Phonetic universals in consonant systems. *Environmental Quality Management*, *2*, 171-176.

Kingston, J. (2007). The phonetics-phonology interface. In: de Lacy, P. (Ed.), *The Cambridge handbook of phonology* (pp. 401-434). Cambridge: Cambridge University Press.

Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus*. Los Angeles, CA: University of California Press.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413-425.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, *125*, 1-15.

Paradis, E. (2012). *Analysis of phylogenetics and evolution with R*. 2nd ed. New York: Springer.

# Author Index

# THE EVOLUTION *of* LANGUAGE

In this proceedings, we collected 195 papers and abstracts from the Joint Conference on Language Evolution (JCoLE), held in Kanazawa, Japan, in September 2022.

Scientific research on language evolution has been making remarkable progress for some decades now, leading to a rapid growth of interdisciplinary research communities with diverse thematic and methodological foci. In 2022, for the first time, we held a joint conference co-organized by Evolang, Protolang, and Evolinguistics. This conference provided an unprecedented opportunity to bring together all the language evolution research communities, enabling a global platform for interdisciplinary discussion.