universität wien

Book of Abstracts

PROTO LANG 9

VIENNA, 15-17/09/25

Co-hosted by

- Department of English and American Studies
- Department of German Studies
- Department of Linguistics
- Department of European and Comparative Literature and Language Studies

Protolang Permanent Organising Committee

Stefan Hartmann (head of the committee) – Slawomir Wacewicz –Przemyslaw Zywiczynski –Francesco Ferretti –Nathalie Gontier – Luke McCrohon –Natalie Uomini –Sverker Johansson –Sylwester Orzechowski – Ines Adornetti – Olga Vasileva –Cedric Boeckx – Evelina Daniela Rodrigues – Piotr Chruszczewski

Protolang 9 Local Organising Committee

Andreas Baumann – Theresa Matzinger – Hannes Fellner – Felix Haiduk

Protolang 9 Program Committee

Aleksandra Ćwiek (Leibniz-Zentrum Allgemeine Sprachwissenschaft) –Andreas Baumann (Universität Vienna) – Angela Stoeger-Horwath (Universität Vienna) – Angelo Damiano Delliponti (Nicolaus Copernicus University in Toruń) – Anna Szala (Oakland University) –Antonio Benítez-Burraco (Universidad de Sevilla) – Bart de Boer (Vrije Universiteit Brussel) –Bart Geurts (Radboud University) – Elizabeth Qing Zhang (Jiangsu Normal University) – Eva Zehentner (University of Zurich) – Felix Haiduk (Austrian Academy of Sciences) – Frederik Hartmann (University of North Texas) – Hannah Brown (University of Victoria) – Hannes A. Fellner (Universität Vienna) – Ilia Afanasev (Universität Vienna) – Ines Adornetti (Roma Tre University) – Irene Boehm (Universität Vienna) – Jonas Nölle (University of Glasgow) – Juan Olvido Perea-Garcia (Universidad de Las Palmas de Gran Canaria) – Kenny Smith (University of Edinburgh) – Limor Raviv (Max-Planck Institute) – Livio Gaeta (University of Turin) – Lukas Nemestothy (Universität Vienna) – Marek Placiński (Nicolaus Copernicus University) – Marta Sibierska (Nicolaus Copernicus University) – Mathilde Josserand (Université Laval) – Michael Pleyer (Nicolaus Copernicus University in Toruń) – Narly Golestani (Universität Vienna) – Niklas Erben Johansson (Lund University) – Piotr Podlipniak (Adam Mickiewicz University) – Przemyslaw Zywiczynski (Nicolaus Copernicus University) – Rie Asano (Teikyo University) – Shiri Lev-Ari (Royal Holloway and Bedford New College) – Simon Kirby (University of Edinburgh, University of Edinburgh) – Slawomir Wacewicz (Nicolaus Copernicus University) – Stefan Hartmann (HHU Düsseldorf) – Svetlana Kuleshova (Nicolaus Copernicus University of Torun) – Theresa Matzinger (Universität Vienna) – Vesta Eleuteri (Universität Vienna) – Yannick Jadoul (University of Roma "La Sapienza")

Administrative support

Nina Hojjat-Apostolidis – Michelle van de Bilt

Protolang 9 Session Chairs

Aleksandra Ćwiek – Angela Stoeger-Horwath – Elizabeth Qing Zhang – Irene Boehm – Limor Raviv – Marta Sibierska – Michael Pleyer – Piotr Podlipniak – Shiri Lev-Ari – Slawomir Wacewicz – Stefan Hartmann – Vesta Eleuteri

Student helpers

Phillippe Colling – Mina Rastoder – David Fleischhacker

Dear Participants,

A very warm welcome to Vienna to the 9th edition of the Protolang conference!

We are pleased that a wide variety of topics are covered in this edition, including species comparisons, corpus analyses, gestures, multimodal aspects, brain imaging, computational modelling and theoretical work, reflecting the width and depth of research in our fields. We are looking forward to the discussions that will emerge and the inspiration they will spark.

The conference comprises:

Three plenary talks by Pritty Patel-Grosz, Tecumseh Fitch, and Anne Kandler

(pp. 4-8)

A workshop on Novel Approaches to Language Evolution Research Using Animal Models (pp. 9-10)

59 talks in parallel sessions (in alphabetical order, pp. 11-223)

We are pleased to invite you to our social dinner on Wednesday to explore the mysteries of the Viennese vineyards. The closing session will happen at dinner.

We want to thank the Departments of English and American Studies, of German Studies, of Linguistics, and of European and Comparative Literature and Language Studies of the University of Vienna for hosting this edition of the conference. We also want to thank the Faculty of Faculty of Philological and Cultural Studies and the Event Management of the University of Vienna.

> Sometimes the unexpected hits us with force and reminds us that what truly matters is the people whose paths we cross in our lives. It is with great sadness that we have to announce the passing of Klaudia Karkowska, a talented young PhD student, who was scheduled to present the talk "What exactly is aligned in the case of alignment in whole-body communication?". To honor her memory, Session II is dedicated to her.

Please find the current schedule here: https://protolang9.univie.ac.at/programme

If you have questions, please contact protolang9@googlegroups.com!

We are committed to creating an inclusive and accessible environment for all participants. If you have any accessibility needs, please let us know so we can ensure your experience is comfortable and engaging.

You can find a **map of the venue** at the end of this book.

PLENARY TALKS

Directive gestures on the primate gestural meaning continuum

Pritty Patel-Grosz[*1]

[*]Corresponding Author: pritty.patel-grosz@iln.uio.no
[1]Department of Linguistics and Scandinavian Studies, University of Oslo, Norway

Research in theoretical semantics has recently expanded its scope from human language to include gestural communication. In parallel, fruitful inquiries at the intersection of primatology and linguistics have given rise to the hypothesis that human and non-human great apes share a common set of directive (= "imperative") gestures. Directive gestures such as STOP or COME-CLOSER pose non-trivial issues for a semantic analysis: we inherit the challenges that pertain to the semantic analysis of imperative utterances (e.g., "Come closer!" in spoken English), while adding a further challenge that stems from the underspecified mapping between a directive body movement and its potential counterparts in human language (e.g., how does the meaning of a STOP gesture compare to the non-equivalent utterances "Be still", "Do not move closer", and "Stop moving closer"). I begin by outlining the problem and surveying the nascent state-of-the-art with regards to a formal modelling of the semantics of directive gestures. Particular attention is given to the multifunctionality of directive gestures, which typically have different effects in different contexts; for example, a non-human ape gesture may communicate "Stop that" in some contexts and "Move away" in others, with similar patterns found in humans. I show how this multifunctionality can be derived from a single, rich abstract lexical entry (amounting to "Not…!" in the case of "Stop that / Move away"); such abstract meanings constitute candidates for universal building blocks of meaning, shared by human and non-human great apes. The emerging framework lays the foundation for expansions of the empirical domain to also include ape gestures (such as ARM-RAISE) found in less studied domains of human communication, such as expressive dancing in a club setting, where the gestures appear to contribute to interpersonal synchronization.

Bridging theory and data in cultural evolution

Anne Kandler*[1]

*Corresponding Author: anne_kandler@eva.mpg.de
[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Understanding the underlying mechanisms that drive change within a system is a fundamental challenge. Often, these mechanisms cannot be directly observed and must instead be inferred from aggregated data. This challenge is not unique to cultural evolution, and significant progress has been made in addressing similar inverse problems in other disciplines. One promising approach to tackle this challenge is generative inference, which employs a generative model—a mathematical representation of the system—to establish causal links between the evolutionary mechanisms in question and observable data, which are then evaluated for statistical consistency. Naturally, the accuracy of this method is influenced by two primary factors: the appropriateness of the generative model in reflecting the key cultural and demographic properties of the system, and the quality of the available data—such as its level of aggregation (e.g., population-level or individual-level data), its sparsity, and its spatial and/or temporal resolution.

In this talk, I will outline how the framework of generative inference can be applied to one of the main questions in cultural evolution: understanding why and how various forms of social learning are used in human populations, both in the present and past. In particular, we will address the issues associated with choosing the appropriate model and imperfect data, highlighting the potential inferential consequences of these issues. First, we revisit a long-standing question: under what circumstances might we expect the ability to learn socially to be favored by selection? Our findings suggest that incorporating human cognitive capacities, such as memory and forgetting processes, can significantly alter predictions regarding the usefulness of social learning. This underscores the importance of considering specific model assumptions when developing frameworks for social learning. Second, using baby name statistics—one of the best-documented cultural datasets—we demonstrate that having "a lot of data" does not necessarily

lead to accurate inference results. Our findings indicate that the presence or absence of rare variants, along with their spread behaviour, may provide a stronger signature of underlying processes than the dynamics of high-frequency variants. Given that this data is often unavailable due to privacy considerations, we illustrate how inference results can vary depending on whether the data is complete or incomplete.

Protolanguages revisited

Tecumseh Fitch[*1]

[*]Corresponding Author: tecumseh.fitch@univie.ac.at
[1]Department of Behavioral and Cognitive Biology, University of Vienna, Austria

Protolanguages are hypothesized intermediate stages in language evolution. Although the term "protolanguage" was introduced by Gordon Hewes in 1973, it was popularized for a particular model of language evolution by Derek Bickerton. For many scientists, Bickerton's model of a protolanguage with words, but without syntax, became the canonical usage of this term, but other very different models of protolanguage have been explored both before and after Bickerton's. I argue that any non-instantaneous model of language evolution needs to have some notion of an intermediate stage or stages, and that the general term "protolanguage" can be used for any of these possible hypothesized intermediates. In previous work I proposed three different plausible hypotheses: musical, lexical, or gestural protolanguage. I suggest that empirical work in language evolution should, whenever possible, test specific predictions of differing models of protolanguage. In this talk I will review models of protolanguage, methods of testing among them, and conclude by suggesting that, given current data, models of language evolution that involve multiple protolanguages are the most plausible option.

WORKSHOP

Novel Approaches to Language Evolution Research Using Animal Models

Organised by Sasha Newar and Tom Jenks

Aim

This workshop offers an opportunity for both human and non-human focused language evolution researchers to exchange knowledge, identify methodological challenges, and explore how cross-disciplinary insights can enhance our collective understanding of language evolution. We aim to show that only by exploring multiple and varied model systems and methods can we truly understand the origins of human language.

Background

The evolution of language is a broad, interdisciplinary field focusing on many different study species. Given its breadth, it is difficult to assimilate relevant key scientific and analytical advancements made in adjacent fields that inform our own research. While we ultimately strive to understand the origins of human language, taking a step back to understand the development of complex communication and vocal learning in nonhuman animals is crucial to understand how we took such a leap in our own abilities. However, as we move away from the human model, it becomes clear that there is no one species that can act as our language out-group, and instead we are left with a collection of remarkable animals with different capabilities, evolutionary histories, and unique methodological requirements to study.

Activity Outline

This workshop will invite speakers to represent their study species (birds, marine mammals, terrestrial mammals), key findings from their research, challenges they face, and novel approaches being used to tackle these problems. We will allocate speaker positions to early career researchers (ECRs) and use their presentations as a platform for further discussion in subsequent break-out sessions. In these break-out sessions principal investigators (PIs) of the speakers will chair activity-based discussions, challenging attendees to evaluate how these novel methods and challenges apply to their own work and systems. The workshop will end with a wider round-table discussion where participants can share new ideas and ways to implement novel techniques into their own research.

Outcomes

This workshop will allow researchers from adjacent fields to connect and form new collaborations to partake in a valuable knowledge exchange. Participants will benefit from engaging with the pioneering work from ECRs and speaking about their own work within the context of the workshop. ECRs will receive valuable feedback on their techniques and the discussion between the brightest minds in language evolution research will lead to the output of novel ideas and collaborations.

Please find more information and the current workshop schedule here: https://verneslab.wordpress.com/protolang-9-workshop

TALKS

Autism and human evolution

Antonio Benítez-Burraco*[1], Ljiljana Progovac[2]

*Corresponding Author: abenitez8@us.es

[1]Department of Spanish, Linguistics & Theory of Literature (Linguistics), University of Seville, Spain

[2]Linguistics Program, Department of English, Wayne State University, Detroit, USA

We consider how both some strengths and some weaknesses exhibited by autism spectrum disorder (ASD) can inform, and be informed by, the considerations of language evolution. We focus on the manifestations of rigidity in ASD, both behavioral and linguistic rigidity, proposing a common (evolutionary) cause: an enhanced striatal function. Moreover, we propose that these features of ASD, even though perceived as negative from the perspective of the typical population, would have been essential for evolving human language.

Here we consider how both some strengths and some weaknesses exhibited by autism spectrum disorder (ASD) can inform, and be informed by, the considerations of language evolution. We focus on the manifestations of rigidity in ASD, both behavioral and linguistic rigidity, and propose a common (evolutionary) cause. Moreover, we propose that these features of ASD, even though perceived as negative from the perspective of the typical population, would have been essential for evolving human language. ASD often exhibits behaviors that, from a typical perspective, can be characterized as rigid, including repetitive, stereotyped behaviors, as well as resistance to changing environments (Kanner, 1943; Bailey et al. 1996; Frith & Happé, 2005; Lord et al., 2020). Behavioral traits associated with ASD also include elevated reactive aggression in some individuals (Hill et al., 2014; Fitzpatrick et al., 2016; Hirota et al., 2020). Importantly for a unified account, some salient characteristics of ASD in the domain of language can equally be characterized as rigid, including heightened sensitivity to the rules of grammar (morpho-syntax), often resulting in hyper-systemizing (e.g. over-regularizing past-tense forms, as in *bring-bringed*) (e.g. Baron-Cohen et al., 2009). Another instance of linguistic rigidity in ASD concerns difficulties with interpreting metaphors (e.g, *belly button*) (Jordan, 2010; Riches et al., 2012; Morsanyi et al., 2020; Lampri et al., 2024), given that

metaphors and other non-literal language require flexibility, i.e. a leap away from the literal/rigid interpretation.

Even though hyper-systemizing yields some atypical language, ASD individuals (and their relatives) often show an increased aptitude for learning rules and for pattern recognition, including those pertaining to language (Baron-Cohen et al., 2009; Ward et al., 2017). Needless to say, the ability to learn rules and patterns is essential for language, and it stands to reason that this ability evolved in humans to support language and other cognitive skills that rely on rules and patterns. ASD is also associated with an enhanced perception of details (e.g. Bor et al. 2007; Happé & Frith 2006; van Leeuwen et al. 2019).

These salient characteristics of rigidity in ASD can be attributed to an enhanced striatal function, as well as to a reduced control of the striatum by selected cortical structures, considering that the striatum is associated with both impulsiveness (including reactive aggression) and with automated, rigid, ritualized responses (Progovac and Benítez-Burraco, 2019; Benítez-Burraco and Progovac, 2021; and references there). The inhibitory function performed by the cortical structures is relevant because typical language processing (and acquisition) results from a delicate balance between the application (and learning) of rules and patterns (i.e. rigidity), and the capacity for suspending such rules when exceptions need to be learnt, and for the purposes of metaphorical extension. Both aspects are crucial for human language and cognition, and we find them enhanced or diminished in certain cognitive conditions such as ASD and schizophrenia. What contributes to a better control of the striatum by the cortical areas of the brain is the high neuronal density of these networks, found altered in conditions like ASD and schizophrenia. A significant enhancement of the connectivity in these networks occurred in relatively recent evolution, implicating FOXP2 and other genes (e.g. Lieberman 2009, Enard et al. 2009, Hillert 2014, Ardila et al. 2016). Such dense connectivity seems to be instrumental not only for exerting control over various physical impulses/reactivity, but also for language processing in general. In this sense, ASD can be characterized as exhibiting a clear strength in one crucial aspect of language evolution, the acquisition and maintenance of the rules of grammar, while at the same time exhibiting a weakness in another, complementary aspect of language evolution, which relies on flexibility in metaphorical extension and in accommodating exceptions. This beneficial effect on learning rules and patterns may also be the reason why this kind of neurodiversity persists in human populations.

References

Ardila A., Bernal B, & Rosselli M. (2016). Why Broca's area damage does not result in classical Broca's aphasia. Front. Human Neurosci. 10, 249. (doi:10.3389/fnhum.2016.00249)

Bailey A, Phillips W, Rutter M (1996) Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. J Child Psychol Psychiatry 37(1):89–126

Baron-Cohen S, Ashwin E, Ashwin C, Tavassoli T, Chakrabarti B. (2009). Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. Phil. Trans. R. Soc. Lond. B 364, 1377-1383.

Benítez-Burraco, Antonio & Ljiljana Progovac (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.

Bor D, Billington J, Baron-Cohen S. (2007). Savant memory for digits in a case of synaesthesia and Asperger syndrome is related to hyperactivity in the lateral prefrontal cortex. Neurocase 13, 311-319.

Enard, W. et al. 2009 A humanized version of FOXP2 affects cortico-basal ganglia circuits in mice. Cell 137, 961–971. (doi:10.1016/j.cell.2009.03.041)

Fitzpatrick SE, Srivorakiat L, Wink LK, Pedapati EV, Erickson CA (2016) Aggression in autism spectrum disorder: presentation and treatment options. Neuropsychiatr Dis Treat 12:1525–1538.

Frith U, Happé F. Autism spectrum disorder. Curr Biol. 2005 Oct 11;15(19):R786-90. doi: 10.1016/j.cub.2005.09.033. PMID: 16213805.

Happé F, Frith U (2006) The weak coherence account: detail-focused cognitive style in autism spectrum disorders. J Autism Dev Disord 36(1):5–25

Hill AP, Zuckerman KE, Hagen AD, Kriz DJ, Duvall SW, van Santen J, Nigg J, Fair D, Fombonne E (2014) Aggressive behavior problems in children with autism spectrum disorders: prevalence and correlates in a large clinical sample. Res Autism Spectr Disord 8:1121–1133.

Hillert, D. (2014). The nature of language: evolution, paradigms and circuits. Berlin, Germany: Springer.

Hirota, T., Deserno, M., & McElroy, E. (2020). The network structure of irritability and aggression in individuals with autism spectrum disorder. Journal of Autism and Developmental Disorders, 50, 1210–1220.

Jordan R (2010) Autism spectrum disorders: an introductory handbook for practitioners. David Fulton Publishers, Oxfordshire.

Kanner L (1943) Autistic disturbances of affective contact. Nervous Child 2(3):217–250.

Lampri S., Peristeri E., Marinis T., & Andreou M. (2024). Metaphor comprehension and production in verbally able children with Autism Spectrum Disorder. Autism Research. 2024 Nov;17(11):2292-2304. (doi: 10.1002/aur.3210)

Lieberman P. (2009). FOXP2 and human cognition. Cell 137, 800–802. (doi:10.1016/j.cell. 2009.05.013)

Lord, K. A., Larson, G., Coppinger, R. P., & Karlsson, E. K. (2020). The history of farm foxes undermines the animal domestication syndrome. Trends in Ecology & Evolution, 35(2), 125–136.

Morsanyi, K., Stamenković, D., & Holyoak, K. J. (2020). Metaphor processing in autism: A systematic review and meta-analysis. Developmental Review 57, Article 100925. (doi.org/10.1016/j.dr.2020.100925)

Progovac, Ljiljana & Antonio Benítez-Burraco (2019). From physical aggression to verbal behavior: Language evolution and self-domestication feedback loop. Frontiers in Psychology 10. 2807.

Riches NG, Loucas T, Baird G, Charman T, Simonoff E (2012) Interpretation of compound nouns by adolescents with specific language impairment and autism spectrum disorders: An investigation of phenotypic overlap. Int J Speech Lang Pathol 14(4):307–317

Van Leeuwen TM, Van Petersen E, Burghoorn F, Dingemanse M, Van Lier R (2019) Autistic traits in synaesthesia: atypical sensory sensitivity and enhanced perception of details. Philos Trans R Soc B 374(1787):20190024

Ward J, Hoadley C, Hughes JE, Smith P, Allison C, Baron-Cohen S, Simner J (2017) Atypical sensory sensitivity as a shared feature between synaesthesia and autism. Sci Rep 7(1):1–9

Automated Training Paradigms in Vocal Learning Bats: A Model Species of Vocal Learning Research

Thomas Jenks[*1,2], Sasha Newar[1,2], Alice Crighton[1,2], Sixue Li[1,2], and Sonja C. Vernes[1,2]

[*]Corresponding Author: tgj1@st-andrews.ac.uk
[1]School of Biology, University of St Andrews, St Andrews, Scotland

While humans show the most extensive display of vocal contextual learning and vocal production learning, it is only by looking at these qualities in other species that we can comprise a comprehensive analysis of these traits which are fundamental to human spoken language. We have developed and tested multiple automated training paradigms using bats as the new model species to assess their intentionality behind vocalisations and their ability to modify the production of their vocalisations.

1. Vocal Learning Species

The origins of human language have been a source of scientific investigations for hundreds of years. Human language appears in over 6,700 different forms (Comrie, 2017) and it affects our cognitive plasticity (Athanasopoulos et al., 2015), concept of time (Boroditsky, 2001), and our ability to express internal concepts to one another (Slobin, 1996). Vocal learning, a key trait underlying the human acquisition of language, refers to two main capabilities. Firstly: *vocal contextual learning* is the ability to learn to produce known sounds in a new context (usage learning) and to understand the new information contained within that known signal in this new context (comprehension learning) (Janik & Slater, 2007). Secondly: *vocal production learning* is the ability to learn to produce new sounds through experience (Janik & Slater, 1997), where the element of learning is the main driver of the change in vocal production. Learning here is described as an individual using memorised information from past experiences to modify their behaviour (Vernes *et al*., 2021). While this is a natural process that we enact each day (learning a new word in our native language, learning a new language, or imitating the sounds around us) vocal production learning is a rare trait. By studying its forms in animal models

which may lack other prelinguistic traits we can form a complete evidence-based analysis of this complex trait and its evolution.

Vocal production learning has only been evidenced in three orders of birds (parrots (e.g. Nottebohm 1972), songbirds (e.g. Kroodsma & Baylis, 1983), and hummingbirds (e.g. Wells & Baptista, 1979) and four clades of non-human mammals (Seals (e.g. Stansbury & Janik, 2019), elephants (e.g. Stoeger *et al*., 2012), cetaceans (e.g. Richards, Wolz & Herman, 1984), and bats (e.g. Knörnschild et al., 2010). While the avian literature has been crucial to our understanding of vocal learning in non-humans they do not utilize a larynx or a mammalian brain structure and so the study of non-human mammals is an important avenue to reveal new insights.

2. Automated vocal learning experiments

The short generation times and ease of husbandry make birds attractive model species and avian research has created the foundation for vocal learning research. However, birds lack the mammalian larynx and brain structure and so bats have emerged as a complementary model for vocal learning research. Training animals for vocal learning experiments can be very time consuming and extracting the relevant information from a species that roosts in large colonies is highly challenging. We have built upon the work of Lattenkamp *et al*. (2018) to produce automated training procedures that allow for the spectro-temporal measurements of individual animals and how they change over time within training paradigms.

In our automated and non-automated paradigms we explore multiple vocal learning related paradigms. Firstly, the vocal usage learning ability of bats to assess the cognitive control bats have over the onset of vocalisations. Following the Nieder & Mooney (2019) framework, we use a discriminative stimulus with no emotional valence, ensure the bats respond within a cognitively relevant timeframe, and ensure that the vocalisations are reliably produced when the cue is present and withheld when the cue is absent. We use this framework to assess the level of control that bats have over the onset of sinusoidally frequency modulated vocalisations in relation to affectively stimulated vocal outbursts. Secondly, we are building from this paradigm to show that bats can listen to a playback and modify their vocalisations to match a playback within this automated paradigm (Lattenkamp *et al*., 2018).

We will show that our automated training paradigm is a flexible and efficient way to train multiple individuals in vocal learning tasks and aim to show that it can be tailored to a multitude of species to aid vocal learning research.

Acknowledgements

Sonja C. Vernes is funded by the UKRI Future Leaders Fellowship and European Research Council Consolidator Grant. Thomas Jenks is funding through the European Research Council funding through Sonja Vernes' grant.

I would like to extend my sincere thanks to Ella Lattenkamp, whose foundational work in establishing an automated training system with bats during her PhD laid the groundwork for this study. It is a privilege to continue building on that legacy. Additionally, thank you to Vincent Janik, whose constant advice and guidance has been invaluable to this project.

References

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N. and Thierry, G. (2015) 'Two Languages, Two Minds: Flexible Cognitive Processing Driven by Language of Operation', Psychological Science, 26(4), pp. 518–526.

Boroditsky, L. (2001) 'Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time', Cognitive Psychology, 43(1), pp. 1–22.

Comrie, B. (2017) 'Languages of the World', in M. Aranoff and J. Rees-Miller (eds) The Handbook of Linguistics. 2nd edn. Blackwell Publishers, pp. 19–42.

Janik, V.M. and Slater, P. (1997) 'Vocal Learning in Mammals', Advances in the Study of Behavior, 26, pp. 59–99.

Knörnschild, M., Nagy, M., Metz, M., Mayer, F. and Von Helversen, O. (2010) 'Complex vocal imitation during ontogeny in a bat', Biology Letters, 6(2), pp. 156–159.

Kroodsma, D.E. and Baylis, J.R. (1983) 'A world survey of evidence for vocal learning in birds', in D.E. Kroodsma and E.H. Miller (eds) Acoustic Communication in Birds. New York: Academic Press, pp. 311–337.

Lattenkamp, E.Z., Vernes, S.C. and Wiegrebe, L. (2018) 'Volitional control of social vocalisations and vocal usage learning in bats', Journal of Experimental Biology, 221(14), pp. 1–8.

Nieder, A. and Mooney, R. (2019) 'The neurobiology of innate, volitional and learned vocalizations in mammals and birds', Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1789).

Nottebohm, F. (1972) 'The Origins of Vocal Learning', The American Naturalist, 106(947), pp. 116–140.

Richards, D.G., Wolz, J.P. and Herman, L. (1984) 'Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, Tursiops truncatus', Journal of Comparative Psychology, 98, pp. 10–28.
Slobin, D.I. (1996) 'From "Thought and Language" to "Language for Speaking"', Rethinking Linguistic Relativity, (July), pp. 70–96.
Stansbury, A.L. and Janik, V.M. (2019) 'Formant Modification through Vocal Production Learning in Gray Seals', Current Biology, 29(13), pp. 2244-2249.e4.
Wells, S. and Baptista, L.F. (1979) 'Displays and Morphology of an Anna x Allen Hummingbird Hybrid', The Wilson Bulletin, 91(4), pp. 524–532.

Balancing similarity and distinctiveness: Competing pressures in the evolution of English monosyllables

Irene Böhm[*1] and Nikolaus Ritt[1]

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

We explore two competing biases in the cultural transmission of words: the need for words (i) to sound alike to be recognizable as words, and (ii) to sound distinctive to avoid confusion. We investigate this dynamic in the development of English monosyllables. Initially, English monosyllables were rare and favored long vowels (XVVC). However, as inflectional endings were lost, monosyllables became much more frequent. We hypothesize that this increased the pressure for phonological differentiation and allowed short XVC shapes to stabilize.

In the cultural transmission of words, two conflicting cognitive biases are at play. On the one hand, there is a pressure for words to have typical sound shapes, so that they can more easily be recognized as words. On the other hand, there is a pressure on words to sound different, so that they are not confused with one another. We explore their dynamics in the evolution of English monosyllables.

Speakers subconsciously track where and how often sounds (co-)occur in the lexicon and use these frequency distributions to identify and process words. Words whose phonotactic shapes are probable – such as those beginning with /bl-/, like *black* or *blue* – are recognized more quickly, learned more easily, and reproduced more accurately (Vitevitch & Luce, 1998, 2016; Storkel, 2001; Goldrick & Larson, 2008, Kelley & Tucker, 2017). This suggests that speakers prefer words with 'canonical' shapes (Wedel, 2006; Blevins, 2006, 2009), and that such words should therefore be selected for in cultural transmission. However, research also demonstrates that words that sound similar to many other words – such as *cat, hat, mat,* or *cap* – are more easily confused, making them harder to identify and retrieve (Vitevitch & Luce, 1999; Vitevitch, 2003; Yates et al., 2004; Goh et al., 2009). It follows that in their cultural transmission, words are shaped by two opposing biases rooted in spoken word processing: the pressure

to conform to common phonotactic shapes and the need to remain distinct enough to avoid confusion.

We explore their interplay in the historical development of English monosyllables. Monosyllables became the canonical word type in English following the decay of the inflectional system, which led to a dramatic increase in their frequency. In earlier periods, when words were still morphologically complex and monosyllabic forms were still in the minority, most monosyllables had long vowels or diphthongs, which selected against short XVC shapes. This is, for instance, evidenced by Early Middle English Open Syllable Lengthening (Minkova & Lefkowitz, 2020), a sound change in which vowels in newly emerging monosyllables where lengthened to conform to the then prototypical XVVC pattern (such as /makə/ > /maːk/ 'make'; Matzinger & Ritt, 2022).

However, as inflectional endings eroded and final schwa was gradually lost, the number of monosyllabic word types increased drastically. We hypothesize that this rise in frequency intensified the pressure for monosyllables to differentiate themselves from one another in terms of their phonological make-up. Given the limited design space of possible monosyllabic forms, XVVC shapes were no longer sufficient to maintain distinctiveness. As a result, previously dispreferred short XVC shapes gained a foothold, eventually stabilized, and even attracted formerly long words (such as /breːd/ > /bred/ 'bread' or /buːk/ > /buk/ 'book'; Ritt, 1997, 2007).

Our study examines the dynamics of expansion of monosyllabic word forms throughout Late Middle English and Early Modern English, as well as in Present-Day English. We use corpus data from the PPCME2 (Kroch & Taylor, 2000), the PPCEME (Korch et al., 2004) the CUBE (Szigetvári & Lindsey, 2013), and the BNC (Davies, 2004) to establish the type and token frequencies of XVC and XVVC shapes in our period of observation and trace how phonotactic majority patterns changed over time.

In our talk, we present the findings of our corpus study and frame our discussion of the competing pressures within the context of how languages exploit available phonotactic design spaces to build words (Tamariz, 2004; Monaghan et al., 2014; Pierrehumbert, 2016; Keogh et al., 2025), how sound shapes can signal morphological structure (Dressler & Dziubalska-Kołaczyk, 2006; Post et al., 2008; Korecky-Kröll, 2014), and ultimately how cognitive processes might drive the selection of specific sound patterns in the cultural evolution of language.

References

Blevins, J. (2006). A theoretical synopsis of evolutionary phonology, *Theoretical Linguistics* 32(2), 117–166.

Blevins, J. (2009). Structure-preserving sound change: A look at unstressed vowel syncope in Austronesian. In A. Adelaar, & A. Pawley (Eds.), *Austronesian historical linguistics and culture history: A festschrift for Bob Blust* (pp. 33–49). Pacific Linguistics.

Davies, M. (2004). *British national corpus from Oxford University Press (BNC)*. https://www.english-corpora.org/bnc/.

Dressler, W. U., & Dziubalska-Kołaczyk, K. (2006). Proposing morphonotactics, *Wiener Linguistische Gazette* 73(1–19), 108–109.

Goh, W. D., Suárez, L., Yap, M. J., & Tan, S. H. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects, *Psychonomic Bulletin & Review* 16, 882-887.

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production, *Cognition* 107(3), 1155–1164.

Kelley, M. C., & Tucker, B. V. (2017). the effects of phonotactic probability on auditory recognition of pseudo-words, *The Journal of the Acoustical Society of America* 141(5), 4038.

Keogh, A., Culbertson, J., & Kirby, S. (2025). *The lexicon adapts to competing communicative pressures: Explaining patterns of word similarity*. PsyArXiv Preprints. https://osf.io/preprints/psyarxiv/zeyua_v1

Korecky-Kröll, K., Dressler, W. U., Freiberger, E. M., Reinisch, E., Mörth, K., & Libben, G. (2014). Morphonotactic and phonotactic processing in German-speaking adults, *Language Sciences*, 46, 48-58.

Kroch, A., & Taylor, A. (2000). *Penn-Helsinki Parsed corpus of Middle English (PPCME2; 2nd ed.)*. University of Pennsylvania. https://www.ling.upenn.edu/ppche/ppche-release-2016/PPCME2-RELEASE-4.

Kroch, A., Santorini, B., & Delfs, L. (2004). Penn-*Helsinki Parsed corpus of Early Modern English (PPCEME)*. University of Pennsylvania. https://www.ling.upenn.edu/ppche/ppche-release-2016/PPCEME-RELEASE-3.

Matzinger, T., & Ritt, N. (2022). Phonotactically probable word shapes represent attractors in the cultural evolution of sound patterns, *Cognitive Linguistics*, 33(2), 415-446.

Minkova, D., & Lefkowitz, M. (2020). Middle English Open Syllable Lengthening (MEOSL) or Middle English Compensatory Lengthening (MECL), *English Language and Linguistics* 25(1), 1–26.

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), 1–12.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic?, *Annual Review of Linguistics* 2, 33–52.

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.

Ritt, N. (1997). Early Middle English vowel shortenings in monosyllables before dentals: a morphologically coded change. In R. Hickey, & S. Puppel (Eds.), *Language history and linguistic modelling* (65-71). Mouton de Gruyter.

Ritt, N. (2007). It seemed like a good idea at the time: why there may not have been any early Modern English shortening before dentals after all. In U. Smit, S. Dollinger, & J. Hüttner (Eds.), *Tracing English through time: Explorations in language variation* (pp. 315-331). Wilhelm Braumüller Universitätsverlag.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development, *Journal of Speech, Language, and Hearing Research* 44(6), 1321–1337.

Szigetvári, P., & Lindsey, G. (2013). *Current British English searchable transcriptions (CUBE).* http://seas.elte.hu/cube/.

Tamariz, M. (2004). *Exploring the adaptive structure of the mental lexicon*. [PhD thesis, University of Edinburgh].

Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the processing of spoken words in English, *Clinical Linguistics & Phonetics* 17, 487–499.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words, *Psychological Science* 9(4), 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition, *Journal of Memory and Language* 40, 374–408.

Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production, *Annual Review of Linguistics* 2(1), 75-94.

Wedel, A. (2006). Exemplar models, evolution and language change, *The Linguistic Review* 23(3), 247–74.

Yates, M., Locker, L. Jr., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception, *Psychonomic Bulletin & Review* 11, 452–57.

Beautiful and hideous words: The effect of cross-linguistic phonesthetics and iconicity on word formation

Niklas Erben Johansson[*1] and Andrey Anikin[2]

[*]Corresponding Author: niklas.erben_johansson@ling.lu.se
[1]Centre for Languages and Literature, Lund University, Lund, Sweden
[2]Division of Cognitive Science, Lund University, Lund, Sweden

Certain speech sounds intrinsically carry meaning, demonstrating that iconicity plays a crucial role in word formation. Some sounds or languages are often also perceived as more beautiful than others, raising the question: Could phonesthetics shape word formation too? We examined phonesthetic judgments across 228 languages, finding that non-tonal languages were perceived as more pleasant. We then investigate tonal languages, exploring links between tone patterns and words with positive or negative valence and how cross-linguistic phonesthetics and iconicity influence the lexicon.

Speech sounds enable us to produce spoken language through double articulation, but they can also carry intrinsic iconic associations to specific meanings, thereby speeding up linguistic transmission. For instance, high and/or rising intonation is often perceived as smaller and more positive than low and/or falling intonation. Similar iconic effects are found across a wide range of semantic domains, influencing words throughout the lexicons of individual languages (Blasi et al., 2016; Erben Johansson et al., 2020; Monaghan & Fletcher, 2019; Sidhu et al., 2021; Winter & Perlman, 2021). This enables iconicity to be used as a crucial strategy for meaning-making and word formation. At the same time, language users frequently perceive certain sounds or languages as more beautiful than others, and specific speech sounds can elicit strong affective arousal (Aryani et al., 2018, 2020), suggesting a potential link between phonesthetic and iconic associations. How, then, might phonesthetics shape the words we use?

We investigated whether there are universal phonesthetic judgments regarding the sound of languages. This study included 2,125 recordings from 228 languages across 43 language families, evaluated by 820 native speakers of English (425 speakers), Chinese (187 speakers), or Semitic languages (200 speakers), who rated how much they liked the sound of each language. These three groups were chosen as they are culturally influential, with different writing systems and profound phonetic differences, yet are well represented in the available pool of

participants. The results showed that recordings of languages perceived as familiar, even when misidentified, and breathy female voices were rated as more pleasant. However, there was little consensus among raters about which languages sounded most beautiful, as personal preferences and perceived resemblance to culturally branded "beautiful" or "ugly" languages significantly influenced judgments. Despite this variability, some population-level phonesthetic preferences were observed. The clearest preference was for non-tonal languages, a trend most pronounced among Chinese-speaking participants—the only tonal participant language in the study. This raises an intriguing question: While lexical tones may be perceived as less pleasing, they are integral to the phonological systems of many languages. How, then, are tones used in tonal languages to convey words with positive or pleasing meanings versus negative or displeasing ones?

In a work-in-progress follow-up study, we collected basic vocabulary items with various positive and negative meanings (e.g., beautiful-ugly, good-bad, correct-wrong, sweet-bitter, happy-sad) from tonal languages representing over 50 language families. The study aims to investigate whether: (a) words with positive valence tend to feature higher and/or more rising tones compared to words with negative valence, aligning with iconic patterns; and (b) words with negative valence exhibit greater tonal complexity, such as more variable pitch contours and tonal contrasts, than words with positive valence. We also explore how tone types can be categorized and quantified to assess how underlying factors influence word formation. Taken together, this presentation examines how our perception of speech, at both the segmental and language levels, facilitates the creation of shortcuts in communication. It also seeks to define the similarities and differences between phonesthetic and iconic associations on a cross-linguistic scale.

References

Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why "piss" is ruder than "pee"? The role of sound in affective meaning making. *PLOS ONE*, *13*(6), e0198430. https://doi.org/10.1371/journal.pone.0198430

Aryani, A., Isbilen, E. S., & Christiansen, M. H. (2020). Affective Arousal Links Sound to Meaning. *Psychological Science*, *31*(8), 978–986. https://doi.org/10.1177/0956797620927967

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

Monaghan, P., & Fletcher, M. (2019). Do sound symbolism effects for written words relate to individual phonemes or to phoneme features? *Language and Cognition*, *11*(2), 235–255. https://doi.org/10.1017/langcog.2019.20

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, *28*(4), 1390–1398. https://doi.org/10.3758/s13423-021-01883-3

Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics*, *6*(1), Article 1. https://doi.org/10.5334/gjgl.1646

Behavioral markers of domestication in vocal learning mammals

Koen de Reus[*1,2,3], Antonio Benítez-Burraco[4], Taylor A. Hersh[1,5], Nikita Groot[6], Megan L. Lambert[7], Katie E. Slocombe[8], Sonja Vernes[6], and Limor Raviv[3,9]

* Corresponding Author: kdereus95@gmail.com

[1] Comparative Bioacoustics Group, MPI for Psycholinguistics, Nijmegen, NL

[2] Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, BE

[3] Donders Centre for Cognition, Radboud University, Nijmegen, NL

[4] Department of Spanish, Linguistics & Theory of Literature, University of Seville, Seville, Spain

[5] Marine Mammal Institute, Oregon State University, Newport, USA

[6] School of Biology, University of St Andrews, St Andrews, UK

[7] Messerli Research Institute, University of Veterinary Medicine Vienna, Vienna, Austria

[8] Department of Psychology, University of York, UK

[9] LEADS group, MPI for Psycholinguistics, Nijmegen, NL

We test the overlap between two relevant phenotypes for language evolution—vocal learning and the domestication syndrome—using an extensive cross-species comparison of 18 domestication traits in vocal learning mammals (humans, African and Asian elephants, humpback whales, bottlenose dolphins, harbor and gray seals, bats, and common marmosets). We find that the hallmark social behavior and cognitive traits of domestication are consistently present in vocal learning mammals, yet there is considerable variation in morphological traits and insufficient data on hormonal markers.

Current work on how language evolved highlights two relevant phenotypes: **Vocal learning (VL)**: the ability to learn new vocalizations or modify existing ones based on auditory experience (Fitch 2000, 2010; Vernes et al., 2021; Jarvis, 2019); and the **Domestication Syndrome (DS)**: a set of behavioral, cognitive, and physiological traits that are typical of domesticated animals (Wilkins et al., 2014; Shilton et al., 2020; Sánchez-Villagra et al., 2016). Notably, both phenotypes are associated with high prosociality, which was closely linked to language emergence (Raviv & Kirby, 2023; Thomas & Kirby, 2018; Benítez-Burraco & Progovac, 2020). But do these two important phenotypes *overlap* in any way? That is, do vocal learning species also show traits that are typically associated with domestication, despite never being domesticated by another

species? While no study to date has examined the relationship between these phenotypes across species, work on marmosets suggests that parental vocal feedback is related to the development of white fur patches (a typical domestication trait; Ghazanfar et al., 2020), and findings from the Bengalese finch hint to a neurobiological link between VL and DS (O'Rourke et al., 2021).

Here, we conducted an extensive cross-species comparison of 18 relevant morphological, hormonal, and behavioral domestication traits in known vocal learning mammals: humans, bats, African and Asian elephants, humpback whales, bottlenose dolphins, harbor seals, and grey seals. We also included common marmosets, for which limited evidence of both vocal learning and domestication traits exist. Results (Figure 1) show that the hallmark social behavior and cognitive traits of DS are consistently present in vocal learning mammals (inc. increased prosociality, cooperation, social tolerance, extensive communication and information sharing, alloparenting, long juvenile periods, and frequent of play behavior that continue well into adulthood). In contrast, there is considerable variation in morphological domestication traits across our study species, potentially due to their distinct ecologies. While there was clear evidence for hormonal markers of domestication in humans and marmosets, there was a notable lack of available data beyond primates. These findings constitute an important first step in assessing the potential evolutionary link between VL and DS, and open up exciting new avenues for comparative work.

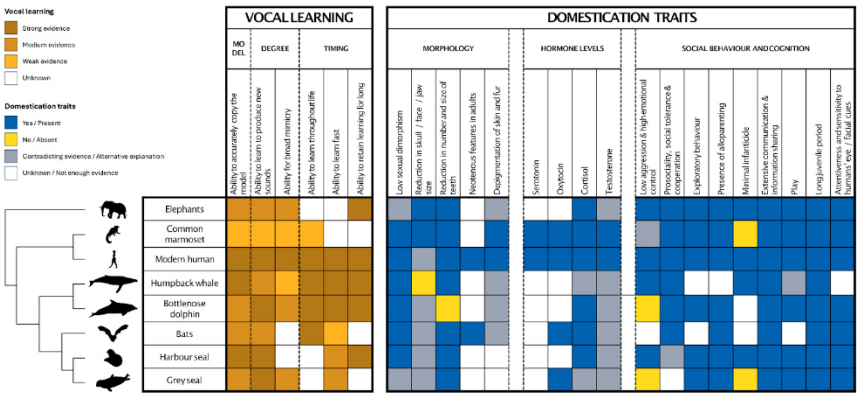


Figure 1. Overview of domestication traits in vocal learning mammals. Each row corresponds to one VL mammal group or species, with evolutionary relationships depicted using a phylogenetic tree. VL features were derived from the dimensions highlighted in Vernes et al. (2021), and color-coded in orange hues based on the strength of available evidence. DS traits were color-coded based on the available evidence in the scientific literature and authors' expertise. For both VL and DS, white cells indicate that no relevant studies were found on the topic, or that the existing evidence was insufficient.

Acknowledgements

KdR was supported by the Fonds Wetenschappelijk Onderzoek project 'Interactive vocal rhythms' G034720N awarded to Bart de Boer. TAH was supported by The Max Planck Gesellschaft and the Oregon Gray Whale License Plate Program at the Marine Mammal Institute (Oregon State University). ABB was supported by MICIU/AEI/10.13039/501100011033 and by ERDF/EU (grant PID2023-147095NB-I00). SCV was supported by a UKRI Future Leaders Fellowship, (MR/T021985/1) and ERC Consolidator Grant (101001702; BATSPEAK). MLL was supported by an FWF Stand-Alone grant (Austrian Science Fund, P34533). We are grateful to Stephanie King for reviewing our bottlenose dolphin assessment.

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.

Fitch, W. T. (2000). The evolution of speech: a comparative review. Trends in Cognitive Sciences, 4(7), 258-267.

Fitch, W. T. (2010). The evolution of language. Cambridge University Press.

Ghazanfar, A. A., Kelly, L. M., Takahashi, D. Y., Winters, S., Terrett, R., & Higham, J. P. (2020). Domestication phenotype linked to vocal behavior in marmoset monkeys. Current Biology, 30(24), 5026-5032.

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. Science, 366(6461), 50-54.

O'Rourke, T., Martins, P. T., Asano, R., Tachibana, R. O., Okanoya, K., & Boeckx, C. (2021). Capturing the effects of domestication on vocal learning complexity. Trends in Cognitive Sciences, 25(6), 462-474.

Raviv, L., & Kirby, S. (2023). *Self-domestication and the Cultural Evolution of Language*. In J. J. Tehrani, J. Kendal, & Rachel Kendal (Eds.), The Oxford Handbook of Cultural Evolution. Oxford Academic.

Sánchez-Villagra, M. R., Geiger, M., & Schneider, R. A. (2016). The taming of the neural crest: a developmental perspective on the origins of morphological covariation in domesticated mammals. Royal Society Open Science, 3(6), 160107.

Shilton, D., Breski, M., Dor, D., & Jablonka, E. (2020). Human social evolution: self-domestication or self-control?. Frontiers in Psychology, 11, 134.

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

Vernes, S. C., Kriengwatana, B. P., Beeck, V. C., Fischer, J., Tyack, P. L., Ten Cate, C., & Janik, V. M. (2021). The multi-dimensional nature of vocal learning. Philosophical Transactions of the Royal Society B: Biological Sciences, 376(1836), 20200236.

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.

Children Are Not the Main Agents of Language Change

Limor Raviv[*1,2], Damian Blasi[3], and Vera Kempe[4]

[*]Corresponding Author: limor.raviv@mpi.nl

[1] LEADS group, Max-Planck-Institute of Psycholinguistics, Nijmegen, The Netherlands

[2] Donders Centre for Cognition (DCC), Radboud University, Nijmegen, the Netherlands

[3] Catalan Institute for Research and Advanced Studies, Barcelona, Spain

[4] Abertay University, Dundee, United Kingdom

We challenge the claim that children are the main agents of language change by showing that (a) innovations that precipitate language change are not unique to children, (b) innovations unique to children (language acquisition errors) are unlikely to diffuse into the community, (c) children's potential role in language emergence does not imply a role in diachronic change, and (d) claims about children driving language change focus on morphosyntax, while phonemic, phonological and lexical change originates in adolescents or adults.

1. The Claim

The claim that children are the main agents of language change has a long tradition in linguistics (e.g. Sweet, 1899). The basic idea is that in the process of language acquisition children produce input-divergent variants (commonly construed as errors) and these innovations then somehow diffuse into the language, either horizontally by spreading into the community (Cournane, 2019) or vertically by transmission to the next generation. In this contribution we aim to discuss the 'holes' in this story.

2. Counterevidence

A prominent argument is that only children produce the kinds of innovations that lead to structural diachronic change. For example, children's linguistic biases, often construed as privileged access to 'Universal Grammar', are responsible for reinterpretation of variable input in a way that can shift languages from one system (e.g. accusative) to another (e.g. ergative) (Chung, 1978). However, such structural input reinterpretations can be shown to arise from statistical learning under cognitive resource limitations (Freudenthal et al., 2006) thereby refuting the idea of children's unique linguistic biases as mechanisms of language change.

Developmental research has attributed the propensity to regularize and overgeneralize, thereby rendering linguistic structure more systematic, to children's cognitive limitations (Hudson Kam & Newport, 2005). However, the kind of unpredictable variation that leads to regularization is non-existent in monolingual language input, and the cognitive constraints that predispose children to regularize can operate in adult learners as well (Perfors, 2012). Thus, we argue that mechanisms responsible for structure-altering innovation, mainly studied in the domain of morphosyntax, are either not empirically validated or not unique to children.

The Achilles' heel of the claim that children drive language change is that no credible mechanism can explain how children's acquisition errors diffuse into the community. Children's recovery from overgeneralization by about 6 years of age (Ambridge et al., 2016) makes it unlikely that such errors can be propagated through vertical transmission. Moreover, human social learning biases favoring frequency or prestige (Henrich, 2001) render children unlikely models for horizontal transmission that would lead communities to alter a highly normative system like language.

Finally, proponents of the view that children drive language change often present evidence from situations of language emergence, e.g. creolization (Bickerton, 1984) or emergence of sign languages (Senghas et al., 2004). Yet children's role in creolization has been called into question due to a lack of structural similarity between creoles (Blasi et al., 2017), and based on evidence documenting a major role of adults in creole formation (Arends, 1993). In sign language formation, e.g. the emergence of Nicaraguan Sign Language, structural innovation has indeed often been introduced by subsequent cohorts of learners, but this is likely not on account of their young age but because they belong to subsequent generations of learners in a transmission chain. Countless iterated learning experiments have demonstrated that the mere process of cultural transmission is bound to introduce structural innovations (Kirby et al., 2008). Moreover, language emergence is qualitatively different from language change as it involves formation of a structured system from an unstructured one rather than change of an already established system. There is no logical reason to assume that mechanisms that explain the former should also explain the latter.

3. Conclusion

Careful scrutiny of the available evidence leads to the conclusion that young children are unlikely to be the main agents of language change. Future research should consider the roles of other age cohorts and the patterns of change in different linguistic domains to gain a better understanding of the mechanisms of language change.

4. References

Ambridge, B., Bidgood, A., Twomey, K. E., Pine, J. M., Rowland, C. F., & Freudenthal, D. (2015). Preemption versus entrenchment: Towards a construction-general solution to the problem of the retreat from verb argument structure overgeneralization. *PloS One, 10*(4), e0123723.

Arends, J. (1993). Towards a gradualist model of creolization. *Atlantic Meets Pacific*, 371–380.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences, 7*(02), 173–188.

Blasi, D. E., Michaelis, S. M., & Haspelmath, M. (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour, 1*(10), 723–729.

Chung, S. (1978). *Case Marking and Grammatical Relations in Polynesian.* University of Texas Press.

Cournane, A. (2019). A developmental view on incrementation in language change. *Theoretical Linguistics, 45*(3–4), 127–150.

Freudenthal, D., Pine, J. M., & Gobet, F. (2006). Modeling the development of children's use of optional infinitives in Dutch and English using MOSAIC. *Cognitive Science, 30*(2), 277–310.

Henrich, J. (2001). Cultural transmission and the diffusion of innovations: adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. *American Anthropologist, 103*(4), 992–1013.

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.

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.

Perfors, A. (2012a). Probability matching vs over-regularization in language: Participant behavior depends on their interpretation of the task. *Proceedings of the Annual Meeting of the Cognitive Science Society, 34*(34).

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.

Sweet, H. (1899). *The practical study of languages: A guide for teachers and learners.* Oxford (reimp. en 1964).

Constraints on number of characters in mimed stories and their implications for scenarios of language origins

Marta Sibierska [*1], Monika Boruta [1], Klaudia Karkowska [1], Bartłomiej Kiljanek[1, 2, 3], Marek Placiński [1], and Antoni Żyndul [1]

[*]Corresponding Author: sibier@umk.pl

[1] Centre for Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland

[2] Institute of Psychology, Nicolaus Copernicus University, Toruń, Poland

[3] Doctoral School of Social Sciences, Nicolaus Copernicus University, Toruń, Poland

Pantomimic scenarios of language origins often underscore the expressive power of pantomime that make it a good medium for conveying semantically advanced content, including stories. On the other hand, previous research has identified constraints on what pantomime is capable of conveying. We designed a comprehension task where we check how many characters can be effectively communicated by means of pantomime, thus determining which types of stories it is able to support.

1. Introduction

Whole-body communication (pantomime *sensu* Żywiczyński et al., 2018) has long been considered a possible precursor of fully-fledged language (e.g., Arbib, 2018; Corballis, 2014; Ferretti et al., 2017, Gärdenfors, 2017; Tomasello, 2008; Zlatev et al., 2020). Pantomimic scenarios of language origins often underscore the expressive power of pantomime that make it a good medium for conveying semantically advanced content, including stories. Indeed, on some accounts (e.g., Ferretti, 2022; Ferretti et al., 2022), sharing stories would have been a paramount function of pantomime before the advent of more language-like forms of communication, which is in line with the stressed significance of story sharing in our prelinguistic prehistory (e.g., Boyd, 2009; Corballis, 2013). On the other hand, previous research has identified constraints on the types of content that pantomime is capable of conveying (e.g., Placiński et al., 2023; Sibierska et al., 2022; Sibierska et al., 2023; Żywiczyński et al., 2021). One important unanswered question concerns the capacity of pantomime to successfully communicate information about characters. Characters are the

semantic units that organise actions and events in stories (Bal, 2017; Jannidis, 2012), their number varying from one or a couple of characters in everyday stories, such as anecdotes or other conversational narratives (cf. e.g., Norrick, 2000), to several or even larger number of characters in conventional forms of storytelling, such as oral literatures (cf. "collective protagonists", e.g. Margolin, 2007). In this study, we examine the limitations on how many characters can be effectively communicated by means of whole-body pantomime, thus determining which types of stories it is able to support.

2. Study

We designed a comprehension task, carried out online, where the participants (N = 70) are asked to watch video clips consisting of front shots of short stories mimed by an actor, in five conditions, corresponding to 1, 2, 3, 4, and 5 characters. We determined the number based on the threshold capacity for information processing and working memory—"the magical number seven" (Miller, 1956)—and the feedback from the actor when faced with the task of communicating stories with up to 7 characters. The actor adopts various strategies to communicate the presence of the characters: from enacting (Müller, 2014; Ortega & Özyürek, 2020) to spatial modulations similar to that used in early stages of the emergence of sign languages (Senghas & Coppola, 2001). After watching a video clip, the participants are asked to answer a comprehension question on the number of characters in the story alongside two "distractor" comprehension questions that also relate to the contents of the stories. Since operationalising comprehension in the case of stories is, in any case, limited (see e.g., Burris & Brown, 2014), we additionally collect data on the self-reported confidence in the answer as well as response times to reflect the perceived difficulty of the task. We assume that with the increasing number of characters in a story, comprehension becomes more challenging.

3. Discussion

Preliminary results from 30 participants support this assumption (final results will be available by the time of the conference). This testifies to the stance that pantomime is curtailed to communicating stories with a certain number of characters. Thinking about our ancestral past, this makes pantomime an optimal medium for story sharing in situations where only a limited number of agents was concerned, possibly in the personal context of small-size communities. In turn, the expansion of relations networks (e.g., moving from esoteric to exoteric niches, see Żywiczyński et al., 2017) and the subsequent need to communicate about more and more people might have brought about a pressure to rely on other means for story sharing.

Acknowledgements

This research has been funded by the National Science Centre of Poland under the agreement UMO 2021/43/D/HS2/01866.

References

Bal, M. (2017). *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press.

Burris, S. E., Brown, D. D. (2014). When all children comprehend: increasing the external validity of narrative comprehension development research. *Frontiers in Psychology* 5: 168. https://doi.org/10.3389/fpsyg.2014.00168.

Ferretti, F. (2022). *Narrative Persuasion. A Cognitive Perspective on Language Evolution*. Springer.

Ferretti, F., Adornetti, I., Chiera, A. (2022). Narrative pantomime: A protolanguage for persuasive communication. *Lingua* 271. https://doi.org/10.1016/j.lingua.2022.103247.

Jannidis, F. (2012). Character. In: P. Hühn et al. (Eds.) *The Living Handbook of Narratology*. Hamburg University. https://www-archiv.fdm.uni-hamburg.de/lhn/node/41.html. DOA: 22 Jul 2024.

Margolin, U. (2007). Character. In: D. Herman (Ed.) *The Cambridge Companion to Narrative*. Cambridge University Press, pp. 66–79. https://doi.org/10.1017/CCOL0521856965.005.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2), pp. 81–97. https://doi.org/10.1037/h0043158.

Müller, C. (2014). Gestural modes of representation as techniques of depiction. In: C. Müller et al. (Eds.) *Body–Language Communication: An International Handbook on Multimodality in Human Interaction*. De Gruyer Mouton, pp. 1687–1701. https://doi.org/10.1515/9783110302028.1687.

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. *Behav Res* 52, pp. 51–67. https://doi.org/10.3758/s13428-019-01204-6.

Placiński, M. et al. (2023). Evolution of Pantomime in Dyadic Interaction. A Motion Capture Study. *Journal of Language Evolution* 8(2), pp. 134–148. https://doi.org/10.1093/jole/lzad010.

Senghas, A., Coppola, M. (2001). Children Creating Language: How Nicaraguan Sign Language Acquired a Spatial Grammar. *Psychological Science* 12(4), pp. 323–328. https://doi.org/10.1111/1467-9280.00359.

Sibierska, M. et al. (2023). Constraints on communicating the order of events in stories through pantomime. *Journal of Language Evolution* 8(1), pp. 18–32. https://doi.org/10.1093/jole/lzad008.

Sibierska, M. et al. (2022). What's in a mime? An exploratory analysis of predictors of communicative success of pantomime. *Interaction Studies* 23(2), pp. 289–321. https://doi.org/10.1075/is.22002.sib.

Żywiczyński, P. et al. (2021). Evolution of conventional communication. A cross-cultural study of pantomimic re-enactments of transitive events. *Language & Communication* 80, pp. 191–203. https://doi.org/10.1016/j.langcom.2021.07.002.

Żywiczyński, P., Wacewicz, S., Sibierska, M. (2018). Defining pantomime for language evolution research. *Topoi* 37, pp. 307–318. https://doi.org/10.1007/s11245-016-9425-9.

Żywiczyński, P., Gontier, N., Wacewicz, S. (2017). The evolution of (proto-)language: Focus on mechanisms. *Language Sciences* 63, pp. 1–11. https://doi.org/10.1016/j.langsci.2017.06.004.

Cultural evolution creates language structure: from humans to humpback whales & beyond

Simon Kirby[*1], Ellen Garland[2], Kazuo Okanoya[3], Jenny Allen[2], Claire Garrigue[5], Emma Carroll[6], Miki Takahasi[7], and Inbal Arnon[8]

[*]Corresponding Author: simon.kirby@ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, UK
[2]School of Biology University of St Andrews, St Andrews, UK
[3]Teikyo Daigaku, Tokyo, Japan
[5]Opération Cétacés, Nouméa, New Caledonia
[6]School of Biological Sciences, University of Auckland, New Zealand
[7]RIKEN Center for Brain Science, Tokyo, Japan
[8]Department of Psychology, Hebrew University, Jerusalem, Israel

All languages have statistically coherent parts whose frequency follows a power law. These properties have not been found elsewhere in nature. We argue that they arise in language via cultural transmission because they facilitate learning. If so, they should be found in other species with culturally transmitted signalling. We illustrate their emergence in humans using an iterated learning paradigm, and then use the same analytic technique to document their presence in whale song and birdsong, also culturally transmitted complex systems.

All known languages are made up of statistically coherent parts whose frequency distribution follows a power law known as a Zipfian distribution (Piantadosi, 2014; Zipf, 1949). Despite the ubiquity of these features their origins are still not understood. In this paper, we argue that they arise because they facilitate learning and therefore emerge through the process of cultural transmission. If so, such properties should also be found in other culturally transmitted communication systems. We use an iterated learning paradigm to illustrate their emergence in human data, and then apply the same analysis to corpora of humpback whale song and birdsong, finding similar structure.

Zipfian distributions facilitate learning in a range of linguistic tasks (e.g., Lavi-Rotbain & Arnon, 2022). We present results from an iterated learning experiment in which non-linguistic sequences evolve as they are transmitted from generation to generation of participants (Cornish et al, 2013). By using

insights from infant speech segmentation (Saffran et al, 1996), we analyse those sequences and observe the emergence of units whose frequency follows a Zipfian distribution over generations (Arnon & Kirby, 2024, Figure 1A). This work makes a prediction that we should find statistically coherent units with a Zipfian distribution whenever sequential behaviour is culturally transmitted, including in other species. However, so far these features have only been found in humans. We apply the same analytic technique to 8 years of humpback whale recordings, another complex culturally transmitted communication system. We find, for the first time in another species, that these characteristic statistical properties are indeed present in whale song (Arnon et al, 2025, Figure 1B). Humpback whales are not the only species with complex culturally transmitted song. Song birds provide another obvious test case, with the advantage that we know a lot more about how their song is learned and produced than we currently know about whales (Takahashi & Okanoya, 2010). Finally, we present ongoing work on the development of the same statistical features in song birds. By doing so, we demonstrate a deep commonality between three species separated by tens of millions of years of evolution but united by having culture.

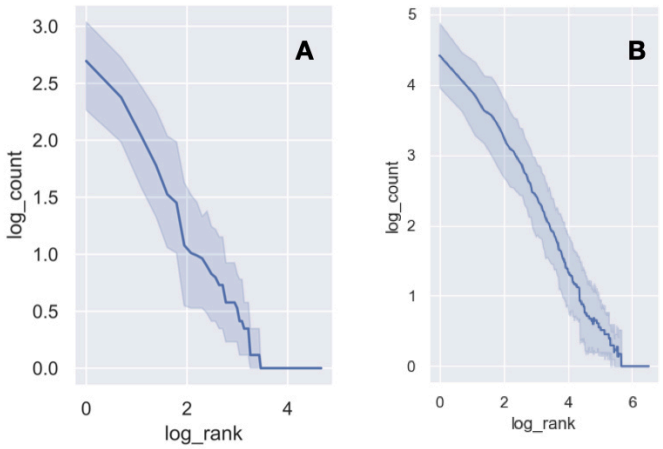


Figure 1. **A**. Zipfian distributions emerge in sub-sequences detected using an infant-inspired segmentation method over generations in an iterated sequence learning task with human participants. The final generation of the iterated learning experiment is shown here. This plot shows the mean and 95% CIs over six chains. The straight line on the log-log plot is diagnostic of a Zipfian distribution. We show that this fit is not an artefact of our method by using a series of baseline analyses. **B**. Sub-sequences detected in humpback whale song using the infant-inspired method exhibit a Zipfian distribution. This plot shows the mean and 95% CIs over eight years of whale song data.

References

Arnon, I., & Kirby, S. (2024). Cultural evolution creates the statistical structure of language. *Scientific Reports*, 14(1), 5255.

Arnon, I., Kirby, S., Allen, J. A., Garrigue, C., Carroll, E. L., & Garland, E. C. (2025). Whale song shows language-like statistical structure. *Science*, 387(6734), 649-653.

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. In P*roceedings of the annual meeting of the cognitive science society* (Vol. 35, No. 35).

Lavi-Rotbain, O., & Arnon, I. (2022). The learnability consequences of Zipfian distributions in language. *Cognition*, 223, 105038.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, 21, 1112-1130.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.

Takahasi, M., & Okanoya, K. (2010). Song learning in wild and domesticated strains of white‐rumped munia, lonchura striata, compared by cross‐fostering procedures: domestication increases song variability by decreasing strain‐specific bias. *Ethology*, 116(5), 396-405.

Zipf, G. (1949). *Human Behavior and the Principle of Least Effort.* New York: Addison-Wesley

.

Dog brains process the meaning of action instruction words

Dorottya S. Rácz[*,1,2], Marianna Boros[1], and Attila Andics[1,3]

[*]Corresponding Author: dorottyaszilviaracz@gmail.com

[1]Neuroethology of Communication Lab, Department of Ethology, Eötvös Loránd University, Budapest, Hungary

[2]Doctoral School of Biology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary

[3]ELTE NAP Canine Brain Research Group

The mechanisms underlying dogs' lexical processing are unclear. In an fMRI study, dogs were presented with action instruction words, phonetically similar pseudowords and dissimilar non-words. Action instruction word processing engaged semantically relevant sensorimotor regions, and the representational geometry of these words reflected semantic similarity. Furthermore, our results suggested coarser-grained auditory word form representations in dogs than in humans. Together, these findings suggest that lexical processing in the dog brain goes beyond word form identification and entails meaning attribution as well.

Dogs are unique in that they have been immersed in the human socio-linguistic niche for thousands of years, have been under selective pressure to communicate efficiently with humans, and are surrounded with speech on a daily basis (Larson et al., 2012; Miklósi, 2007; Thalmann et al., 2013). Thus, speech has become a relevant part of their natural environment, making them an ideal model to test whether brain specialisations for speech processing are unique to humans or can emerge as a rapid evolutionary adaptation in response to exposure to speech.

Dogs are not only sensitive to nonverbal human communicative cues (Bray et al., 2021; Miklósi et al., 2005; Byosiere et al., 2022), but they often show adequate behaviour to words directed to them (Fukuzawa et al., 2005; Ramos & Mills, 2019). In humans, spoken word understanding requires both auditory word form identification and meaning attribution to that form. But how these contribute to lexical processing in dogs is not known (Gábor et al., 2020). Auditory word form identification should take place in the auditory cortex and, if human-like, should not tolerate speech sound changes (Flemming, 2004), as understanding that even single speech sound changes may result in a different word can be crucial for building large vocabularies. Meaning attribution, in contrast, may also engage non-auditory, semantically relevant brain regions, as seen in humans (Carota et al., 2017; Pulvermüller, 2013, 2018). To seek evidence for both processes in the dog brain, we presented dogs (N=21) with (1) instruction words for actions requiring locomotion or not (L+, L-), (2) phonetically similar pseudowords (generated from words by altering a single speech sound) and (3) dissimilar non-words (Magyari et al. 2020) in an fMRI experiment. First, we searched for lexically responsive brain regions by contrasting words to non-words. We

expected that, if understood, action instruction words will engage not only auditory but also relevant motor cortical regions. Next, we assessed brain responses to pseudowords in all lexically responsive brain regions to probe the rigidity of auditory word form representations. Finally, we looked for representational similarities comparing response patterns to instruction words for actions with vs. without locomotion.

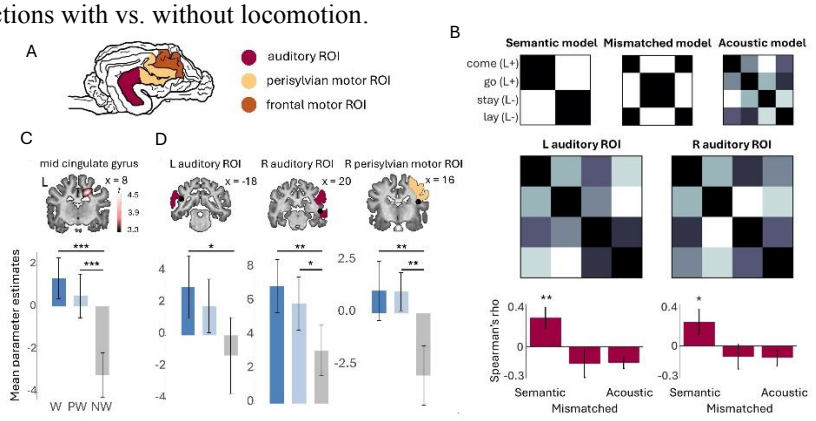


Figure 1. Results. (A) Regions of interest. (B) Lexical effects in a whole-brain GLM. Word > non-word contrast overlaid on a template dog brain. (C) Lexical effects in the ROIs. Peaks obtained in small volume-corrected GLMs shown as black dots. Bar graphs represent parameter estimates in the peaks. (D) Representational similarity analysis for semantic relatedness. Top panels: model RDMs (semantic: semantic similarity-matched; mismatched: semantic similarity-mismatched; acoustic). Middle panels: subject-averaged neural RDMs for word stimuli in the auditory ROIs. Bottom panels: Spearman's rank correlation between model RDMs and neural RDMs. Error bars represent SE. *p<0.05, ** p<0.01, *** p<0.001. L = left, R = right. W = word, PW = pseudoword, NW = non-word.

Lexical effects were found in semantically relevant motor (Walker et al., 1962), and motor control regions (Hoffstaedter et al., 2014; Paus, 2001, Fig.1A-C), indicating meaning attribution. Additionally, instruction words requiring more similar action preparation were represented as more similar in the dog auditory cortex (Fig.1D) suggesting that, beyond recognizing known speech sound sequences (Boros, Magyari et al., 2021; Gábor et al., 2020), the auditory cortex is also involved in the semantic analysis of the word forms similarly than in the human brain (Damera et al., 2023; DeWitt & Rauschecker, 2012). The processing of pseudowords, however, did not differ from that of words in either of the lexically responsive brain regions (Fig.1B-C) indicating that auditory word form representations in the dog brain are coarser-grained than in humans, which may be related to dogs' limited vocabulary. Thus, this study provides the first fMRI evidence that the dog brain goes beyond auditory word form identification and entails meaning attribution as well. Despite the obvious differences in dog and human linguistic capacities, these results raise the possibility that the cognitive and neural architecture underlying linguistic meaning processing is similar, with mental representations organized along semantic similarities in both species, challenging human uniqueness accounts.

Acknowledgements

This project was funded by the Hungarian Academy of Sciences via a grant to the MTA-ELTE 'Lendület' Neuroethology of Communication Research Group (grant no. LP2017–13/2017) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 950159). A.A. was supported by the National Brain Programme 3.0 of the Hungarian Academy of Sciences (NAP2022-I-3/2022). D.S.R. was supported by the EKÖP-24 University Excellence Scholarship Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund (EKÖP-24-3-I-ELTE-58).

References

Boros, M., Magyari, L., Török, D., Bozsik, A., Deme, A., Andics, A., 2021. Neural processes underlying statistical learning for speech segmentation in dogs. Curr. Biol. 31, 5512-5521.e5. https://doi.org/10.1016/j.cub.2021.10.017

Bray, E. E., Gnanadesikan, G. E., Horschler, D. J., Levy, K. M., Kennedy, B. S., Famula, T. R., & MacLean, E. L. (2021). Early-emerging and highly heritable sensitivity to human communication in dogs. *Current Biology*, *31*(14), 3132-3136.e5. https://doi.org/10.1016/j.cub.2021.04.055

Byosiere, S. E., Mundry, R., Range, F., & Virányi, Z. (2022). Selective responding to human ostensive communication is an early developing capacity of domestic dogs. *Developmental Science*. https://doi.org/10.1111/desc.13361

Carota, F., Kriegeskorte, N., Nili, H., Pulvermüller, F., 2017. Representational Similarity Mapping of Distributional Semantics in Left Inferior Frontal, Middle Temporal, and Motor Cortex. Cereb. Cortex 27, 294–309. https://doi.org/10.1093/cercor/bhw379

Carota, F., Nili, H., Pulvermüller, F., Kriegeskorte, N., 2021. Distinct fronto-temporal substrates of distributional and taxonomic similarity among words: evidence from RSA of BOLD signals. Neuroimage 224, 117408. https://doi.org/https://doi.org/10.1016/j.neuroimage.2020.117408

Damera, S., Chang, L., Nikolov, P., Mattei, J., Banerjee, S., Glezer, L., Cox, P., Jiang, X., Rauschecker, J., Riesenhuber, M., 2023. Evidence for a Spoken Word Lexicon in the Auditory Ventral Stream. Neurobiol. Lang. 4, 1–40. https://doi.org/10.1162/nol_a_00108

DeWitt, I., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. 109, E505–E514. https://doi.org/10.1073/pnas.1113427109

Flemming, E., 2004. Contrast and perceptual distinctiveness, in: Hayes, B., Kirchner, R., Steriade, D. (Eds.), The Phonetic Bases of Markedness. Cambridge University Press, Cambridge, UK.

Fukuzawa, M., Mills, D.S., Cooper, J.J., 2005. The effect of human command

phonetic characteristics on auditory cognition in dogs (Canis familiaris). J. Comp. Psychol. https://doi.org/10.1037/0735-7036.119.1.117
Gábor, A., Gácsi, M., Szabó, D., Miklósi, Á., Kubinyi, E., Andics, A., 2020. Multilevel fMRI adaptation for spoken word processing in the awake dog brain. Sci. Rep. 10, 11968. https://doi.org/10.1038/s41598-020-68821-6
Hoffstaedter, F., Grefkes, C., Caspers, S., Roski, C., Palomero-Gallagher, N., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2014. The role of anterior midcingulate cortex in cognitive motor control. Hum. Brain Mapp. 35, 2741–2753. https://doi.org/https://doi.org/10.1002/hbm.22363
Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y. W., Peters, J., Stahl, P. W., Piper, P. J., Lingaas, F., Fredholm, M., Comstock, K. E., Modiano, J. F., Schelling, C., Agoulnik, A. I., Leegwater, P. A., Dobney, K., Vigne, J. D., Vilà, C., Andersson, L., & Lindblad-Toh, K. (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proceedings of the National Academy of Sciences of the United States of America, 109(23), 8878–8883. https://doi.org/10.1073/pnas.1203005109
Magyari, L., Huszár, Z., Turzó, A., Andics, A., 2020. Event-related potentials reveal limited readiness to access phonetic details during word processing in dogs. R. Soc. Open Sci. 7, 200851. https://doi.org/10.1098/rsos.200851
Mallikarjun, A., Shroads, E., Newman, R., 2021. The role of linguistic experience in the development of the consonant bias. Anim. Cogn. 24. https://doi.org/10.1007/s10071-020-01436-6
Miklósi, Á. (2007). Dog Behaviour, Evolution, and Cognition. Oxford University Press.
Miklósi, Á., Pongrácz, P., Lakatos, G., Topál, J., & Csányi, V. (2005). A comparative study of the use of visual communicative signals in interactions between dogs (Canis familiaris) and humans and cats (Felis catus) and humans. Journal of Comparative Psychology, 119(2), 179–186. https://doi.org/10.1037/0735-7036.119.2.179
Obleser, J., Zimmermann, J., Van Meter, J., Rauschecker, J.P., 2007. Multiple Stages of Auditory Speech Perception Reflected in Event-Related fMRI. Cereb. Cortex 17, 2251–2257. https://doi.org/10.1093/cercor/bhl133
Paus, T., 2001. Primate anterior cingulate cortex: Where motor control, drive and cognition interface. Nat. Rev. Neurosci. 2, 417–424. https://doi.org/10.1038/35077500
Peelle, J.E., Johnsrude, I., Davis, M.H., 2010. Hierarchical processing for speech in human auditory cortex and beyond. Front. Hum. Neurosci. 4.
Pulvermüller, F., 2018. Neural reuse of action perception circuits for language, concepts and communication. Prog. Neurobiol. 160, 1–44. https://doi.org/https://doi.org/10.1016/j.pneurobio.2017.07.001
Pulvermüller, F., 2013. How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. Trends Cogn. Sci. 17, 458–

470. https://doi.org/https://doi.org/10.1016/j.tics.2013.06.004
Ramos, D., Mills, D.S., 2019. Limitations in the learning of verbal content by dogs during the training of OBJECT and ACTION commands. J. Vet. Behav. Clin. Appl. Res. 31, 92–99. https://doi.org/10.1016/j.jveb.2019.03.011
Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., & Sawyer, S. K. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*, *342*(November), 871–874.
Walker, S.M., Baggish, M.S., Katz, D.B., Noss, C.G., Robert, B.C., Truong, X.T., 1962. A second motor area in the cerebral cortex of the dog. J. Neurophysiol. 25, 607–620. https://doi.org/10.1152/jn.1962.25.5.607

Do Facial Signals Support Language Emergence During Communication? An Experimental Approach

Loïs Dona[*1,2], Aslı Özyürek[1], Judith Holler,[1,3], Marieke Woensdregt[3], and Limor Raviv[2]

[*]Corresponding Author: lois.dona@mpi.nl

[1]Multimodal Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

[2]LEADS, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

[3]Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, Netherlands

This study explores the role of facial signals in language emergence, an area that is relatively underexplored, by manipulating facial visibility in a communication game. Analyzing communicative success, convergence, stability and response times, we find that while facial signals can influence communication, they are not essential for building a successful, stable and converged communication system. In future research, we will examine potential compensatory strategies participants may have used to adapt to the absence of facial signals.

The primary mode of (early) human communication is face-to-face interaction, involving multimodal and meta-communicative signals such as facial signals (Vigliocco et al., 2014). Apart from expressing emotion, it is now widely recognized that facial signals can support communication in existing languages by e.g. facilitating the prediction of upcoming utterances, repair, backchanneling and signalling of uncertainty (Emmendorfer & Holler, 2025; Hömke et al., 2025; Micklos & Woensdregt, 2023; Nölle et al., 2023; Bavelas & Chovil, 2018; Hömke et al., 2018; Swerts & Krahmer, 2005). These kinds of processes likely also play a role in the pragmatic inference relevant to coordinating on novel linguistic variants (Healey et al., 2007; Micklos et al., 2018; Roberts et al., 2018). However, communication experiments that simulate the process of coordinating and converging on novel linguistic variants often overlook the role of such meta-communicative facial signals. Even more strikingly, many of them are not conducted face-to-face. Contrary to a lack of experimental work, there do exist theories about the role of the face in language evolution, which largely highlight the affordance of the face facilitating the

transition between a gesture-dominant and a vocalization-dominant system (see Wacewicz et al. (2016) for an overview). However, the current work aims at investigating the role of the face in the emergence of conventionalized symbol systems more broadly, and how facial signals can serve as a meta-communicative tool to help establish common ground in interaction.

To address this gap, we make use of a communication game paradigm to explore the role of facial signals in language emergence. We manipulate facial visibility in a between-subjects design (34 dyads total) yielding two conditions: face visible and face invisible. Dyads repeatedly communicate about twelve shapes (adapted from Macuch Silva et al. (2020)) while being video and audio recorded, with roles of producer and guesser switching every trial: the producer invents and writes a novel label for the target shape, and the guesser guesses what the target shape is from an array of shapes upon seeing the label. Writing (or speaking) in existing language was prohibited. Participants played the game, using the same shapes, over four consecutive rounds.

We measured (1) communicative success, capturing accuracy of a guess, (2) convergence, capturing how similarly participants in a dyad label a shape, and (3) stability, capturing how similarly an individual labels a shape over time. We have also conducted an analysis of guessing and labeling times. Results provide insight into how facial signals affect the communication and coordination involved in building a novel communication system, contributing to a deeper understanding of the face's role in language emergence.

Participants who do not see each other lack access to facial signals that support pragmatic inference, for example by signalling feedback (Hömke et al., 2018, 2025), but also confidence regarding the label or guess (Nölle et al., 2023; Swerts & Krahmer, 2005). We predicted that a lack of this information would hinder communicative success and convergence. However, we also acknowledge the possibility that participants adapt to the lack of facial signals in ways that facilitate communication, such as creating more motivated labels (G. Roberts et al., 2015; Tamariz et al., 2018) or meta-communicative vocalizations.

Results show communicative success, convergence and stability increasing significantly over rounds ($p < .001$), but no significant main or interaction effect of condition. However, the face invisible condition showed significantly more variance for convergence ($p < .001$): pairs who could see each other converged more uniformly, while convergence was more varied across pairs who could not. Furthermore, the analysis of guessing and labeling times revealed that both were slower when participants interacted face-to-face (guessing: $p < .001$, labeling: $p < .05$). Moreover, this effect interacts with communicative success and convergence for guessing time ($p < .001$): trials with incorrect guesses or the use of less converged on labels were much slower when the face was visible compared to invisible.

A highly plausible explanation for the longer response times is the

increased processing load that comes with participants attending to facial signals, especially when they are less confident in their guess. In addition, the greater variation in convergence in the face invisible condition suggests that emerging communication systems are more susceptible to variation when facial signals are unavailable. Despite these observations, the overall similar levels of communicative success and convergence between conditions suggest that facial signals are not indispensable for success or convergence. We speculate that participants can adaptively overcome the absence of facial signals through alternative means, such as using more iconicity or meta-communicative vocalizations. Taken together, these findings suggest that while facial signals can influence early communication, they are not essential for building a successful communication system, at least in the written medium. To get a more nuanced understanding of how the face influences communicative success and convergence, we will explore the role of meta-communicative vocalizations and iconicity of the labels as potential compensatory strategies, while also examining the role of facial expressivity in future research.

Acknowledgements
We would like to thank the authors of Macuch Silva et al. (2020) for using part of their stimuli. In addition, we would like to thank Noam Siegelman for sharing the set of alien figures that were used as stimuli for practice trials.

References

Bavelas, J., & Chovil, N. (2018). Some pragmatic functions of conversational facial gestures. *Gesture*, *17*(1), 98–127. https://doi.org/10.1075/gest.00012.bav

Emmendorfer, A. K., & Holler, J. (2025). Facial signals shape predictions about the nature of upcoming conversational responses. *Scientific Reports*, *15*(1), 1381. https://doi.org/10.1038/s41598-025-85192-y

Healey, P. G. T., Swoboda, N., Umata, I., & King, J. (2007). Graphical Language Games: Interactional Constraints on Representational Form. *Cognitive Science*, *31*(2), 285–309. https://doi.org/10.1080/15326900701221363

Hömke, P., Holler, J., & Levinson, S. C. (2018). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLOS ONE*, *13*(12), e0208030. https://doi.org/10.1371/journal.pone.0208030

Hömke, P., Levinson, S. C., Emmendorfer, A. K., & Holler, J. (2025). Eyebrow movements as signals of communicative problems in human face-to-face interaction. *Royal Society Open Science*, *12*(3), 241632. https://doi.org/10.1098/rsos.241632

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.

https://doi.org/10.1098/rsos.182056
Micklos, A., Silva, V. M., & Fay, N. (2018). The prevalence of repair in studies of language evolution. *Proceedings of the 12th International Conference on the Evolution of Language (Evolang12)*. The Evolution of Language. Proceedings of the 12th International Conference on the Evolution of Language (Evolang12). https://doi.org/10.12775/3991-1.075
Micklos, A., & Woensdregt, M. (2023). Cognitive and Interactive Mechanisms for Mutual Understanding in Conversation. In *Oxford Research Encyclopedia of Communication*. https://doi.org/10.1093/acrefore/9780190228613.013.134
Nölle, J., Wu, Y., Arias, P., Yang, Y., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2023). *Multimodal markers of confidence and doubt: Inferring Feeling of Knowing from facial movements*.
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. https://doi.org/10.1016/j.cognition.2015.04.001
Roberts, S., Micklos, A., Sulik, J., & Little, H. (2018). Innovation, selection and the emergence of transparent signals in interaction. *The Evolution of Language: Proceedings of the 12th International Conference (EVOLANGXII)*.
Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94. https://doi.org/10.1016/j.jml.2005.02.003
Tamariz, M., Roberts, S. G., Martínez, J. I., & Santiago, J. (2018). The Interactive Origin of Iconicity. *Cognitive Science*, *42*(1), 334–349. https://doi.org/10.1111/cogs.12497
Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130292. https://doi.org/10.1098/rstb.2013.0292
Wacewicz, S., Żywiczyński, P., & Orzechowski, S. (2016). Visible movements of the orofacial area: Evidence for gestural or multimodal theories of language evolution? *Gesture*, *15*(2), 250–282. https://doi.org/10.1075/gest.15.2.05wac

Evolution of brain connectivity for language: Arcuate fascicle termination in the middle temporal gyrus across great apes

Gereon Ohagen[1,#], Yannick Becker[1,#,*], Cornelius Eichner[1], Michael Paquette[1], Christian Bock[2], Carsten Jäger[3,4], Tobias Gräßle[5], EBC Consortium, Philipp Gunz[6], Roman M. Wittig[7,8,9], Catherine Crockford[7,8,9], Angela D. Friederici[1], Alfred Anwander[1]

[#]These authors contributed equally
*Corresponding Author: beckery@cbs.mpg.de

[1]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[2]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[3]Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[4]Paul Flechsig Institute – Center of Neuropathology and Brain Research, Medical Faculty, University of Leipzig, Germany
[5]Ecology and Emergence of Zoonotic Diseases, Helmholtz Institute for One Health, Helmholtz Centre for Infection Research, Greifswald, Germany
[6]Department of Human Origins, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[7]Evolution of Brain Connectivity Project, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[8]Institute for Cognitive Sciences Marc Jeannerod, UMR CNRS 5229, University Claude Bernard Lyon 1, Bron, France
[9]Taï Chimpanzee Project, CSRS, Abidjan, Ivory Coast

The evolution of the neural basis of human language remains an open question. The arcuate fascicle (AF) is the main language fiber connection in the human brain. Previously, AF termination in the middle temporal gyrus (MTG) was considered unique to humans. Using white matter tractography, we identify AF-MTG terminations in all great ape species. This suggests an evolutionary onset 12–15 million years ago, potentially providing an important neural scaffold for the evolution of language.

Comparing the brains of different primate species is crucial to understanding primate – including human – evolution. Together with grey matter structures, the white matter connectivity is the fundamental component of brain organisation. However, a study comparing white matter connectivity across all ape species, our closest primate relatives, is lacking. One of the most distinctive features of the human species is language. The main fibre tract for human language in the brain is the arcuate fascicle (AF). It connects crucial frontal and temporal language regions in the human brain, notably Broca's area and Wernicke's area (Friederici et al., 2017). The AF is typically left lateralised in the human species and displays two distinct terminations in the temporal lobe: one in the superior temporal gyrus

(STG) and another in the middle temporal gyrus (MTG). These crucial features are previously thought to have arisen *de novo* in the human species (e.g., Eichert et al., 2019; Roumazeilles et al., 2020; Sierpowska et al., 2022; Friederici & Becker, 2025). In contrast, while other nonhuman primate species display an AF connection, they generally exhibit only an STG termination, lacking both the MTG termination and the characteristic left lateralisation (e.g., Eichert et al., 2019; Roumazeilles et al., 2020; Sierpowska et al., 2022; Friederici & Becker, 2025). Recent research on chimpanzee brains has provided systematic evidence for a weak MTG termination in both captive and wild subject (Becker et al., 2025). Interestingly brain growth between wild and captive chimpanzees seem not to differ (Cofran, 2019). To trace back the evolutionary onset of this distinct brain feature for language and to study potential species differences, we compared the connectivity strength of the AF-STG and AF-MTG in all great ape species (humans, bonobos, chimpanzees, gorillas, orangutans) using high-resolution MRI tractography. We conducted probabilistic FSL based tractography on 15 subjects (3 brains per species: 12 post-mortem brains from captive non-human primates (bonobo, chimpanzee, orangutan, gorilla) and 3 *in vivo* brains from human primates). Using the comparative approach, we aim to pinpoint the evolutionary timeframe for the emergence of a precursor of the neural substrate of language.

A termination of the arcuate fascicle (AF) in the middle temporal gyrus (MTG) was found in each hemisphere of each individual in each of the five great ape species.

Therefore, we propose that the MTG termination of the AF did not emerge *de novo* in the human lineage. Instead, our findings suggest that the human configuration evolved through strengthening an MTG connection of the AF that already existed at least 12-16 million years ago, when our last common ancestor with orangutans lived. This preexisting pathway may have provided an important neural scaffold for the evolution of human language.

Future studies using behavioural data from the same subjects, when they were alive, will now need to investigate what behaviours the AF-MTG brain connection has evolved to support in apes and whether such behaviours could be precursors to syntax and semantics, as it is the case in human for the language faculty.

References

Becker, Y., Eichner, C., Paquette, M., Girard-Buttoz, C., Bock, C., Jäger, C., Deschner, T., EBC Consortium, Gunz, P., Wittig, R., Crockford, C., Friederici, A., Anwander, A. (in press*).* Long arcuate fascicle in wild and captive chimpanzees as a possible structural precursor of the language network. *Nature Communications*. https://doi.org/10.1038/s41467-025-59254-8

Cofran, Z. (2018). Brain size growth in wild and captive chimpanzees (Pan troglodytes). American Journal of Primatology, 80(7), e22876

Eichert, N., Verhagen, L., Folloni, D., Jbabdi, S., Khrapitchev, A. A., Sibson, N. R., Mantini, D., Sallet, J., & Mars, R. B. (2019). What is special about the human arcuate fasciculus? Lateralization, projections, and expansion. *Cortex*, *118*, 107–115. https://doi.org/10.1016/j.cortex.2018.05.005

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. The MIT Press.

Friederici, A. D., & Becker, Y. (2024). The core language network separated from other networks during primate evolution. *Nature Reviews Neuroscience*, 1–2. https://doi.org/10.1038/s41583-024-00897-9

Roumazeilles, L., Eichert, N., Bryant, K. L., Folloni, D., Sallet, J., Vijayakumar, S., Foxley, S., Tendler, B. C., Jbabdi, S., Reveley, C., Verhagen, L., Dershowitz, L. B., Guthrie, M., Flach, E., Miller, K. L., & Mars, R. B. (2020). Longitudinal connections and the organization of the temporal cortex in macaques, great apes, and humans. *PLOS Biology*, *18*(7), e3000810. https://doi.org/10.1371/journal.pbio.3000810

Sierpowska, J., Bryant, K. L., Janssen, N., Blazquez Freches, G., Römkens, M., Mangnus, M., Mars, R. B., & Piai, V. (2022). Comparing human and chimpanzee temporal lobe neuroanatomy reveals modifications to human language hubs beyond the frontotemporal arcuate fasciculus. *Proceedings of the National Academy of Sciences*, *119*(28), e2118295119. https://doi.org/10.1073/pnas.2118295119

Experiments in language evolution: inferences from closed- vs open-ended semantic space paradigms

Svetlana Kuleshova[*1,2], Stefan Hartmann[3], Michael Pleyer[1], Marta Sibierska[1], Johan Blomberg[4], Przemysław Żywiczyński[1], Aleksandra Ćwiek[1,5], Marek Placiński[1], Sławomir Wacewicz[1]

[*]Corresponding Author: 40010189@parisnanterre.fr

[1]Nicolaus Copernicus University in Toruń, Center for Language Evolution Studies, Fosa Staromiejska 3, 87-100 Toruń, Poland.

[2]ArScAn-Équipe AnTET (UMR 7041), CNRS, Université Paris Nanterre, 21 Allée de l'Université, 92000, Nanterre, France.

[3]Heinrich Heine University Düsseldorf, Faculty of Arts and Humanities, Universitätsstraße 1, 40225 Düsseldorf, Germany.

[4]Stockholm University, Department of Swedish and Multilingualism. Universitetsvägen 10, 106 91 Stockholm.

[5]Leibniz-Centre General Linguistics, Berlin, 10719, Germany

We challenge the validity of the multiple-choice paradigm used in many communication experiments. We replicated two experiments and replaced the original multiple-choice design with free-text answers. Our results indicate that participants performed worse compared to the original studies. The post hoc analysis showed that often relevant semantic domains were correctly identified, but full congruency with the target concepts was rare. We discuss how the results of each paradigm can be mapped onto a larger picture of language evolution.

What inferences about reality can we make from linguistic experiments? By definition, controlled experiments do not represent reality in all its complexity, and thus, it is important to be cautious when drawing conclusions. As one example, some recent experimental studies have made strong claims concluding that humans can *understand* improvised or interspecies forms of communication, specifically, novel non-linguistic vocalizations (Ćwiek et al., 2021) and gestures in non-human great apes (Graham and Hobaiter, 2023). Here, we follow up on these studies, altering one crucial aspect: Both original studies used a multiple-choice paradigm, whereby participants were constrained to choose one answer from a small set of alternatives; that is, for each gesture or vocalisation they were

offered a set of potential interpretations and asked to guess the one best reflecting its intended meaning. We expected that replacing the closed-ended, multiple-choice paradigm with an open-ended, free-text paradigm would lead to different results and thus different conclusions about human capacities to 'understand' ape gestures or novel vocalizations.

To test our hypothesis, we replicated the studies by Ćwiek et al. and Graham & Hobaiter, but with the participants freely typing in their answer (i.e., their interpretation of the meaning of an ape gesture or novel vocalisation) rather than selecting it from a set of predefined choices. To assess the proximity of the participants' answers with the target concepts, we used three complementary approaches: (i) binary assessment of correctness (yes/no), (ii) manual coding on a four-point scale, (iii) cosine similarity between the participants' responses and the target concepts using ConceptNet. For (i) and (ii), participants' answers were coded by two raters on a four-point Likert scale (transformed into a binary system for (i)), showing moderate-to-high inter-rater agreement. All mismatches between coders were discussed in the team and resolved. For (iii), the proximity between the words in the responses (heads of phrases in the case of multi-word answers) and the word denoting the target concept in ConceptNet (Speer et al. 2017) were assessed automatically. Across all approaches, our results show that participants performed qualitatively worse compared to the original studies, with a vast majority of responses showing no resemblance whatsoever to the target concept according to the manual coding procedure. In the ConceptNet analysis, the responses were found to be conceptually fairly close to the target concepts in many cases (e.g., the same semantic domain, as in 'sleep' vs 'snore'), but hardly any response was fully congruent with the target concept.

Apart from revisiting the question to what extent participants "understand" novel vocalizations or ape gestures, our data give us the opportunity to evaluate different possibilities of interpreting participants' free-text answers: manual coding with different levels of granularity on the one hand, and automatic similarity detection using NLP resources on the other. The three measures we used carve out different semantic landscapes, shaping how we perceive communicative transparency. A perfect-match approach treats meaning as a binary on/off switch, limiting "understanding" to exact hits while discarding near-misses as failures. The Likert-scale approach expands this space, recognizing perceived shades of similarity. The ConceptNet-based cosine similarity approach stretches the landscape even further, mapping words onto a gradient of conceptual distances, where even broad associations register as meaningful connections. Because these methods define similarity differently, they potentially reveal different facets of how novel signals convey meaning.

We thus conclude that different setups and analysis methods give us a nuanced perspective on the role of semantic space in semiotic experiments. These various paths to data collection and data treatment can inform a nuanced discussion of how experimental results can inform the evolution of today's languages.

Acknowledgements

PZ was supported by the National Science Centre of Poland (NCN) under grant agreement UMO-2023/49/B/HS2/0222.
MP was supported by the National Science Centre of Poland (NCN) under grant agreement UMO-2024/53/B/HS2/01366.
MS was supported by the National Science Centre of Poland (NCN) under the agreement UMO 2021/43/D/HS2/01866.
AĆ was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant number CW10/1-1, within SPP 2392 "Visual Communication (ViCom)"

References

Ćwiek, A., et al. (2021). Novel vocalizations are understood across cultures. *Sci Rep 11*, 10108 .

Graham, K. E., & Hobaiter, C. (2023). Towards a great ape dictionary: Inexperienced humans understand common nonhuman ape gestures. *PLoS Biology, 21(1)*, e3001939.

Speer, Robyn, Joshua Chin & Catherine Havasi. (2017). ConceptNet 5.5: An open multilingual graph of general knowledge. http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972.

Exploiting epidemiological contact data to examine the role of age structure in linguistic diffusion

Andreas Baumann[*1,3], Frederik Hartmann[4], and Yllka Velaj[2,3]

[*]Corresponding Author: andreas.baumann@univie.ac.at
[1]Faculty of Philological and Cultural Studies, University of Vienna, Vienna, Austria
[2]Faculty of Computer Science, University of Vienna, Vienna, Austria
[3]Research Network Data Science, University of Vienna, Vienna, Austria
[4]College of Engineering, University of North Texas, Denton, USA

Age is a much-discussed factor in language change. In this paper, we investigate the role that age structure plays in the diffusion of linguistic innovations. For this purpose, we analyze predictions from an age-structured population dynamic model of linguistic diffusion. Next to demographic and psycholinguistic data, our model is informed by age-structured contact data. We provide further support for the hypothesis that adolescents play a central role in the diffusion of innovations.

To what extent language change is associated with age was the subject of abundant linguistic research. For one, there is much discussion about the respective roles that children and adults play in language change. While the extreme position that language change is exclusively driven by the process of language acquisition in children is likely not plausible (e.g., Niyogi 2006), there is some evidence that early acquisition has stabilizing effects (Monaghan & Roberts 2019; Elsherif et al. 2023). In addition to that, socio-linguistic research has revealed insights into which age class might function as a motor for linguistic innovations (Labov 1972).

In this contribution, we investigate the impact that different age classes have on the diffusion of linguistic innovations with the help of ordinary differential equation models of linguistic diffusion informed by demographic data as well as data about contact events collected for epidemiological research (cf. Hoang et al. 2019). Our results indicate that, from a population-dynamic point of view, adolescents rather than children are a driving force in the spread of linguistic innovations (Kerswill 1996).

We use an age-structured version of susceptible-infected-susceptible (SIS) models (Anderson & May 1991) of diffusion. SIS models, which essentially model sigmoid growth (Denison 2003) of an innovation in a population, have

been already applied in linguistic research in (Nowak 2000; Baumann 2024). We integrate eight disjoint age classes (Fogarty et al. 2019). Individuals from two age classes (or indeed the same age class) can meet in which case transmission of the linguistic innovation will take place at a certain rate. Individuals can also cease using the innovation. Ageing is not possible in the current version of the model so that it only captures short-term developments. Under the assumption that transmission is strong enough, the model displays a non-trivial stable and attracting population dynamic equilibrium.

To inform the model, i.e., to provide estimates of the model parameters, we use diverse datasets, restricting our analysis to the German speaker population. Rates of contact among age classes are derived from a large-scale contact pattern dataset provided by Mossong et al. (2008; see Figure 1a), discretized into eight age classes. Information about demographic age structure (i.e., relative size per age class) is taken from the UN (https://data.un.org/; Figure 1b, insert). In a second step, we also examine differential transmission rates for which we use (age specific) lexical-decision times for German from DeveL (Schröter & Schroeder 2017) as weights to enhance cognitive plausibility.

Given the model parameters, we use the model to predict trajectories of the prevalence of a (hypothetical) linguistic innovation, one for each age class. It is shown that irrespective of which age class an innovation originates from it is adolescents who display particularly rapid diffusion of the innovation (Figure 1b) due to their relatively high connectivity. A sensitivity analysis of age-related parameters is also provided to learn about the relative importance of each age class. In addition to our findings regarding the role of age, we take our model to represent a reasonable baseline for investigating more complex socio-pragmatic phenomena (like preferred accommodation among peers).

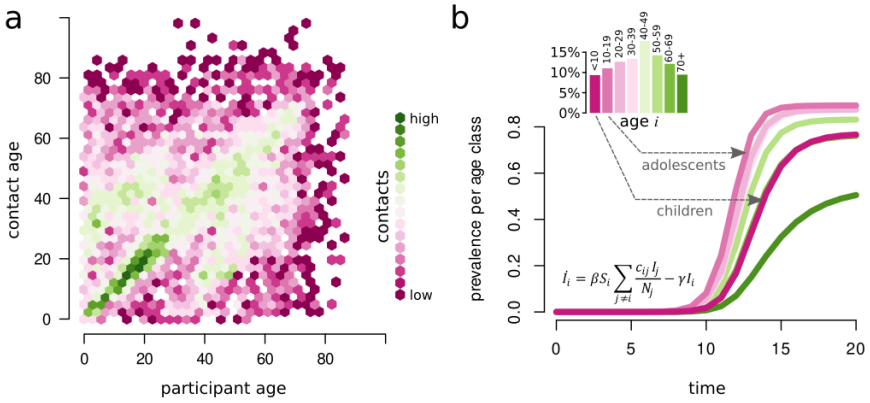


Figure 1. (a) Physical-contact distribution in Germany as provided by Mossong et al. (2008). (b) Trajectories of a hypothetical linguistic innovation for each age group derived from the age-structured model (here without age-specific transmission rates).

References

Anderson, R. M., & May, R. M. (1991). Infectious diseases of humans: dynamics and control. Oxford university press.

Baumann, A. (2024). Lexical innovations are rarely passed on during one's lifetime: epidemiological perspectives on estimating the basic reproductive ratio of words. PLOS One. https://doi.org/10.1371/journal.pone.0312336

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

Elsherif, M. M., Preece, E., & Catling, J. C. (2023). Age-of-acquisition effects: A literature review. Journal of Experimental Psychology: Learning, Memory, and Cognition, 49(5), 812.

Fogarty, L., Creanza, N., & Feldman, M. W. (2019). The life history of learning: Demographic structure changes cultural outcomes. PLoS Computational Biology, 15(4), e1006821.

Hoang, T., Coletti, P., Melegaro, A., Wallinga, J., Grijalva, C. G., Edmunds, J. W., ... & Hens, N. (2019). A systematic review of social contact surveys to inform transmission models of close-contact infections. Epidemiology, 30(5), 723-736.

Kerswill, P. (1996). Children, adolescents, and language change. Language Variation & Change, 8(2).

Labov, W. (1972). Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

Monaghan, P., & Roberts, S. G. (2019). Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. Cognition, 186, 147-158.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., ... & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine, 5(3), e74.

Niyogi, P. (2006). The computational nature of language learning and evolution (Vol. 43). Cambridge, MA: MIT press.

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.

Schröter, P., & Schroeder, S. (2017). The Developmental Lexicon Project: A behavioral database to investigate visual word recognition across the lifespan. Behavior Research Methods, 49, 2183-2203.

Exploring the Neural Architecture of Communication in Non-Human Great Apes

M. Sinha[*1], M. Mohebbi[1], A. Murphy[1], W.D. Hopkins[2] , C.C. Sherwood[3], D. Stout[4], E.E. Hecht[1]

[*]mirasinha@g.harvard.edu
[1]Human Evolutionary Biology, Harvard, Cambridge, USA [2]MD Anderson Cancer Ctr., Univ. of Texas, USA [3]Anthropology, George Washington Univ., Washington DC, USA
[4]Emory Univ., Atlanta, USA

This study investigates the neural basis of communication by examining neuroimaging data across multiple great ape species. The dataset includes a sample of language-trained individuals and enculturated non-language-trained apes, enabling a comparison of species-specific neural structures and the effects of language exposure. By analyzing key communication pathways, such as the arcuate fasciculus and ventral pathways, we uncover patterns of pathway asymmetry and dorsal-ventral ratios that reflect evolutionary constraints and plasticity, providing insights into the evolution of language.

1.1. Introduction

Our closest relatives provide valuable insights into the evolutionary roots of language. Training and enculturation programs using hand signs and lexigrams have explored the extent of non-human great ape communication abilities. Behavioral experiments showed that most apes could utilize and combine communicative signals in a meaningful manner. However, their active vocabulary and syntactical orders were substantially less complex than humans (Lyn, 2012). These apparent differences may be explained by neurobiological changes in humans compared to apes. Humans have larger and more specialized brain regions dedicated to language processing, as supported by neuroimaging studies showing significant structural differences between humans and non-human primates (Friederici, 2009; Schenker et al., 2010; Rilling et al., 2008; Eichert et al., 2019). An important question is to what extent these neural features are a result of innate biological processes versus exposure to a language-rich environment. This study explores species differences in neural networks involved in communication, as well as experience-dependent plasticity related to language training.

1.2. Method

We utilized a neuroimaging dataset that includes MRI scans from multiple great ape species — 37 humans, 67 in vivo chimpanzees, 18 *ex vivo*

chimpanzees, and 9 gorillas — offering an opportunity to explore innate species-specific brain organization and the effects of language training. The 5 language-trained apes in our study were trained using lexigrams or American Sign Language over a span of more than 20 years. Their training methods varied, with some apes receiving instruction through operant conditioning, while others were trained in socio-cultural environments. In addition, we examined age-, sex-, and species-matched non-language-trained apes. Ex vivo brains were collected opportunistically after the animals' natural deaths. This allowed for direct comparison of brain structure across different training conditions within each species.

We focused on white matter regions involved in language processing, particularly the arcuate fasciculus (AF) and ventral pathways, which are critical for language processing in humans (Catani & Mesulam, 2008; Hickok & Poeppel, 2007). Our aim was to examine species differences in brain structure and connectivity of these pathways, as well as any experience-dependent changes in language-trained apes.

1.3. Results

Initial analyses of brain structure variability related to communication in great apes indicate interspecies differences. Notably, humans exhibit a marked leftward asymmetry in the arcuate fasciculus (AF), while non-human apes demonstrate more bilateral or less pronounced asymmetry. Similarly, we observed distinctions in the dorsal-ventral pathway ratio, with humans showing a higher dorsal-ventral ratio compared to other great apes. Compared to chimpanzees, humans display a significantly more robust AF (normalized by a control tract; left AF: $p < 0.0001$; right AF: $p < 0.0001$) and greater leftward asymmetry ($p = 0.0008$). In contrast, chimpanzees exhibit a relatively larger ventral tract when similarly normalized (left VP: $p < 0.0001$; right VP: $p < 0.0001$); see Fig 1.

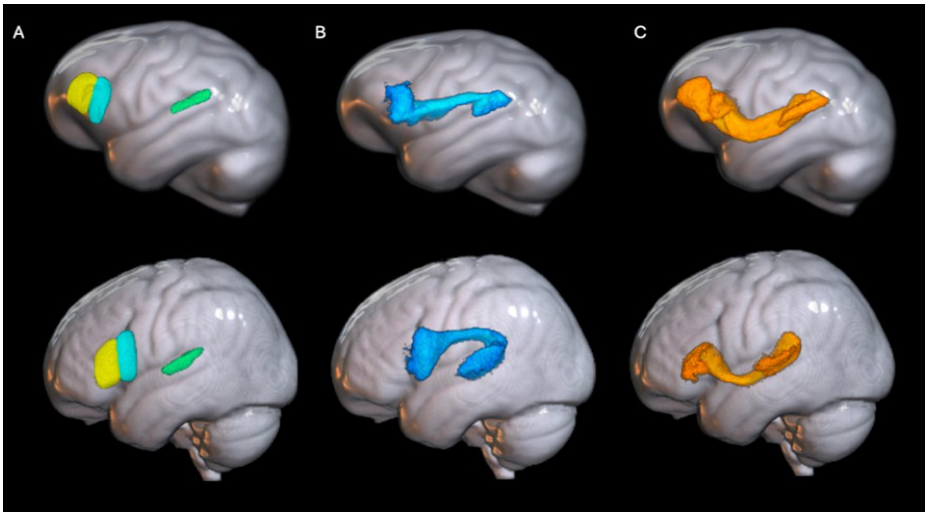


Fig 1. Chimpanzee (top) and human (bottom) regions of interest and tractography results in Juna (Chimpanzee) and MNI (Human) template space (Vickery et al., 2021). A: Regions of Interest: Inferior Frontal Gyrus (IFG; yellow = aIFG, blue = pIFG) and Planum Temporale (PT). B: Tractography of the reconstructed Arcuate Fasciculus (Blue). C: Tractography of the reconstructed Ventral Pathway (Orange).

Within language-trained apes, variability in AF asymmetry was observed. For example, in the chimpanzee sample, Lana's AF, unlike Panzee's, exhibited a shift toward a human-like leftward asymmetry pattern (Lana's AF: 2nd percentile; value = -1.66; chimpanzee mean = 0.068, SD = 0.613, z = -2.82). However, there was considerable variability among language trained apes, possibly due to differences in training methods and the age at first language exposure.

1.4. Conclusion

Together, these findings illustrate the evolutionary divergence in communication neural networks between humans and other great apes. Humans demonstrate greater AF specialization and higher dorsal/ventral tract ratios, highlighting these regions' role in advanced language processing. Despite the intra-group variation introduced by the diverse enculturated environments of language-trained apes, this study provides the first opportunity to investigate how ape brain organization is modified by experience related to language training. By examining both species differences and experience effects, these insights may improve our understanding of neural communication circuits and provide perspectives on language evolution in *Homo sapiens*, emphasizing how biological and cultural factors shape the neural substrates of communication.

References

Lyn, H. (2012). 19 Apes and the Evolution of Language: Taking Stock of 40 Years of Research. *The Oxford handbook of comparative evolutionary psychology*, 356.

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends in cognitive sciences*, *13*(4), 175-181.

Schenker, N. M., Hopkins, W. D., Spocter, M. A., Garrison, A. R., Stimpson, C. D., Erwin, J. M., ... & Sherwood, C. C. (2010). Broca's area homologue in chimpanzees (Pan troglodytes): probabilistic mapping, asymmetry, and comparison to humans. *Cerebral Cortex*, *20*(3), 730-742.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature neuroscience*, *11*(4), 426-428.

Eichert, N., Verhagen, L., Folloni, D., Jbabdi, S., Khrapitchev, A. A., Sibson, N. R., ... & Mars, R. B. (2019). What is special about the human arcuate fasciculus? Lateralization, projections, and expansion. *cortex*, *118*, 107-115.

Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *cortex*, *44*(8), 953-961.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, *8*(5), 393-402.

Vickery, S., Hopkins, W. D., Sherwood, C. C., Schapiro, S. J., Latzman, R. D., Caspers, S., ... & Hoffstaedter, F. (2020). Chimpanzee brain morphometry utilizing standardized MRI preprocessing and macroanatomical annotations. *Elife*, *9*, e60136.

Sherwood, C. C., Cranfield, M. R., Mehlman, P. T., Lilly, A. A., Garbe, J. A. L., Whittier, C. A., ... & Hof, P. R. (2004). Brain structure variation in great apes, with attention to the mountain gorilla (Gorilla beringei beringei). *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *63*(3), 149-164.

Expressibility ratings as a predictor of communicative success in referential games

Aleksandra Ćwiek[*1,2], Susanne Fuchs[1], Wim Pouw[3,4], and Šárka Kadavá[1,4,5]

[*]Corresponding Author: cwiek@leibniz-zas.de
[1]Leibniz-Centre General Linguistics, Berlin, Germany
[2]Center for Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland
[3]Department of Cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, The Netherlands
[4]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[5]University of Göttingen, Göttingen, Germany

This study validates expressibility ratings as efficient proxies for communicative success in referential games. We demonstrate strong correlations between subjective expressibility judgments and objective guessability across modalities (gesture, vocalization, combined), measured as perfect match and cosine similarity. Results show that gesture and combined modalities perform better than vocalization. Feedback modestly improved perfect matches but not semantic similarity. The approach offers methodological advantages for cross-linguistic investigations into modality affordances and provides an efficient tool for studying emergent communication systems.

Referential games, i.e., tasks in which people have to negotiate a meaning using novel communicative signals, provide evidence about language at its dawn, specifically regarding the communicative success of novel signals across various modality conditions. However, such experiments require substantial financial, time, and human resources and, due to the multitude of degrees of freedom, their results may be biased (Fay et al., 2014; Macuch Silva et al., 2020). This study validates an efficient approach using expressibility ratings—subjective judgments of how easily concepts can be communicated in different modalities—as predictors of actual communicative success. It also examines the effects of modality choice and the perceiver's feedback on communication effectiveness.

We tested whether expressibility ratings collected online could predict guessability in laboratory experiments. Dutch speakers (N = 248) rated how

expressible 207 concepts would be using only gestures, only vocalizations, or a combination of both. Subsequently, 142 participants (71 dyads) were recruited to take part in a referential game with 84 concepts from the original set, with performers conveying concepts in assigned modalities and guessers attempting to identify them.

Using Bayesian hierarchical models, we analyzed the relationship between expressibility ratings and guessability, measured as both binary correctness and semantic similarity (cosine similarity using ConceptNet embeddings; Speer et al., 2017). Results strongly supported our primary hypothesis that expressibility ratings predict guessability ($\beta = 7.30$, 95% CrI [6.80, 7.81] for binary correctness; $\beta = 1.02$, 95% CrI [0.93, 1.11] for semantic similarity). Higher expressibility was associated with fewer correction attempts ($\beta = -1.59$, 95% CrI [–1.79, –1.38]). Regarding modality effects, multimodal and gesture-only conditions yielded comparable guessability (binary model: ~46% and ~45% correct), while vocalization-only performed significantly worse (binary model: ~12% correct), partially supporting our hypothesized hierarchy: combined > gesture > vocal. The introduction of the perceiver's feedback modestly improved binary correctness (~30% to ~34%) but not semantic similarity, suggesting feedback primarily helps participants pinpoint exact labels but not necessarily to the guesser being more aligned with the performer's intention.

Our findings validate expressibility ratings as efficient proxies for communicative success. The strong correlation between subjective judgments and real-life guessability indicates accurate awareness of modality affordances, contrary to concerns about how overconfident people are about their own judgment (Kahneman & Klein, 2009). The modality success patterns align with research suggesting that visual modality (either only in gesture or combined with vocalization) provides an advantage in novel communication, i.e., communication without existing conventions (Fay et al., 2014). However, we acknowledge that modality affordances vary by semantic domain (Kadavá et al., 2024).

Our self-rating approach, therefore, presents a cost- and time-efficient tool that can complement, or, in some specific cases, replace costly production experiments. For instance, it enables large cross-linguistic investigations or can inform stimulus selection for referential communication experiments, helping researchers avoid inadvertent biases. The expressibility ratings database (Ćwiek et al., in prep.) represents a resource for testing theoretical and methodological assumptions (Ćwiek et al., in prep.) related to modality differences in the emergence of symbolic communication systems.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – CW 10/1-1 (AĆ), FU 791/9-1 (SF and ŠK), and PO 2841/1-1 (WP) – SPP 2392 "Visual Communication (ViCom)."

References

Ćwiek, A., Fuchs, S., Pouw, W., & Kadavá, Š. (in prep.). *Self-reported expressibility predicts communicative success: Open dataset, validation, and simulation.*

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a communication system from scratch: Gesture beats vocalization hands down. *Frontiers in Psychology, 5*.

Kadavá, Š., Ćwiek, A., Fuchs, S., & Pouw, W. (2024). What do we mean when we say gestures are more expressive than vocalizations? An experimental and simulation study. *Proceedings of CogSci 2024, 46*, 2308–2315.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515–526.

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.

Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. *Proc. of the AAAI Conference on Artificial Intelligence, 31*(1).

Has Child-Directed Speech Changed Over the Past Six Decades?

Vera Kempe*[1], C. Donnan Gravelle[2], Stina Perschke[1], Glenn P. Williams[3], and Sonja Schaeffler[4]

*Corresponding Author: v.kempe@abertay.ac.uk
[1]Department of Sociological & Psychological Sciences, Abertay University, Dundee, UK
[2]The Graduate Center, City University of New York, USA
[3]Department of Psychology, Northumbria University, Newcastle upon Tyne, UK
[4]Clinical Audiology, Speech and Language Research Centre, Queen Margaret University, Edinburgh, UK

Unlike its cross-cultural variation, historical change in child-directed speech (CDS) has not been studied. We analyze its prosodic features in German child-directed broadcasts (CDB) from 1959 to present and compare them to adult-directed broadcasts (ADB) comprising weather forecasts. The CDB series originated in East Germany and continued post-reunification (1990). While CD pitch variation did not change over time, slower articulation rates in CDS emerged only after 1990, possibly reflecting a socio-cultural shift towards greater child-centeredness. We conclude that CDS may have undergone historical change, potentially accelerated by historical events.

1. Introduction

Large-scale cross-cultural studies have identified prosodic features of CDS such as higher pitch, increased pitch variation, and slower articulation rates as universal (e.g. Cox et al., 2023). However, most research on CDS has focused on the past 30 years, offering only a historical snapshot (Kempe et al., 2024). As human behavior changes over time (Muthukrishna et al., 2021), a full understanding of CDS requires examining potential historical change. This study leverages established similarities between CDB and CDS (Zhang & Gu, 2023) to explore change over time in CDS. We analyzed a corpus of German CDB dating from 1959 to 2023, which originated in East Germany and continued after German reunification in 1990, subsequently covering the entire German transmission area. Given that East Germany adhered to a collectivist ideology that emphasized directive childrearing, while post-reunification Germany followed Western liberal-democratic values, we hypothesized that a sociocultural shift toward greater child-centeredness may be observable in an amplification of certain CDS characteristics in post-1990 German CDB.

2. Method

We analyzed pitch variation and articulation rate of 103 story-telling monologues of the German children's program *Unser Sandmännchen* (1959–present),

comprising 35 actors addressing an imaginary child audience. Deutsches Rundfunk Archiv (DRA) provided access to pre-1990 broadcasts; post-1990 broadcasts were obtained from regional broadcasters (RBB, NDR) and YouTube. We compared these to ADB comprising 192 weather forecasts presented by 90 different speakers, covering the same period. Weather forecasts were chosen for their temporal continuity in content.

3. Results

Linear mixed-effect regression discontinuity models were fitted to model change over time. For pitch variation, the model estimated higher marginal means for female (M=43.4 Hz, SE=3.63) compared to male speakers (M=29.1 Hz, SE=1.19; β=0.53, z=6.47, p<.001), and for CDB (M=49.4 Hz, SE=3.92) compared to ADB (M=29.7 Hz, SE=1.44; β=0.53, z=5.04, p<.001). These findings align with expectations about differences in pitch variation between genders (Pépiot, 2014) and between ADS vs. CDS (Cox et al., 2023). Trend estimates showed an increase in pitch variation over time in ADB (β=0.10, SE=0.04, z=2.66, p=0.008), likely indicative of a shift towards infotainment, but no change over time in CDB (β=0.03, SE=0.06, z=-0.57, p=0.57), attesting to the stability of higher pitch variation in this register. For articulation rate, a Register x Discontinuity interaction (β=1.28, z=3.53, p<.001) indicated that pre-1990, there was no difference between CDB (M=4.44 syll/sec, SE=0.17) and ADB (M=4.18 syll/sec, SE=0.16; t(275)=1.03, p=0.31). However, post-1990, the model estimated slower articulation rates in CDB (M=3.53 syll/sec, SE=0.27) than ADB (M=4.51 syll/sec, SE=0.100; t(275)=3.29, p=0.001). Thus, the slower articulation rate commonly found in recent studies on CDS was not evident in pre-1990 East German CDB, and only emerged post-1990.

4. Discussion

Our findings suggest that some aspects of CDS may have varied over historical time in response to changing social norms, and that sometimes such change may reflect rapid shifts due to historical events. Specifically, greater child-centeredness, associated with a sociocultural shift after German reunification, manifested itself in slower CDS, an indicator of speech accommodation that takes account of children's needs, but not in changes in pitch variation, a likely more universal indicator of conventionalized positive affect expression. Our study raises questions about the universality of CDS features and emphasizes the value of taking a historical approach to the study of CDS.

5. References

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2023). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour, 7*(1), 114-133.

Kempe, V., Ota, M., & Schaeffler, S. (2024). Does child-directed speech facilitate language development in all domains? A study space analysis of the existing evidence. *Developmental Review, 72,* 101121.

Muthukrishna, M., Henrich, J., & Slingerland, E. (2021). Psychology as a historical science. *Annual Review of Psychology, 72*(1), 717-749.

Pépiot, E. (2014). Male and female speech: A study of mean F0, F0 range, phonation type, and speech rate in Parisian French and American English speakers. *Speech Prosody 7*, 305–309.

Zhang, Y., & Gu, Y. (2023). A recipient design in multimodal language on TV: A comparison of child-directed and adult-directed broadcasting. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Hierarchical structure of kinship semantics derives from efficient communication principles

Maisy Hallam[*1], Simon Kirby[1], Fiona M. Jordan[2], and Kenny Smith[1]

[*]Corresponding Author: maisy.hallam@ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, United Kingdom
[2]Department of Anthropology and Archaeology, University of Bristol, United Kingdom

Kin terms, the set of words we use to denote family relationships, map variably to meanings. But are there common semantic cues to category membership? Under the assumption that wordforms map to meanings in a compressible, structure-preserving manner, we measured the relative contribution of six semantic features to category membership. We uncovered a cross-linguistic hierarchy in which features are more reliable cues to shared category membership, suggesting that kinship semantics are constrained by pressures associated with efficient communication.

Systems of kinship terminology differ considerably in which semantic distinctions they encode. For instance, an English speaker distinguishes their *brother* from their *sister* on the basis of gender, while an Indonesian speaker distinguishes *kakak* 'older sibling' from *adik* 'younger sibling', on the basis of their relative age. Here, we explore whether there are constraints determining which semantic features are likely to cue shared category membership.

Despite an enormous amount of cross-linguistic variation, kinship terminologies are constrained by pressures associated with efficient communication (Kemp & Regier, 2012; Kemp et al., 2018). For a particular level of communicative precision, kin terms are maximally compressible, and for a particular level of compression they are maximally precise. Lexicons achieve compression by mapping similar meanings to similar wordforms — by being structure-preserving, or **topographic** (Brighton & Kirby, 2006). For instance, in English, the wordform *uncle* maps to two concepts: 'mother's brother' and 'father's brother'. Words can also be topographic at the morphosyntactic level, i.e. **compositional**: the English kin term *grandson* combines the morphemes *son*, meaning male child, and *grand-*, indicating a two-generation distance from the speaker.

Using these insights about the expected relationship between meaning and form, we explore whether kin terms exhibit universal tendencies in which semantic features tend to predict shared form. We start with the assumption that kinship terminologies are indeed topographic and use this to uncover their underlying semantic structure. We measured the extent to which a variety of seman-

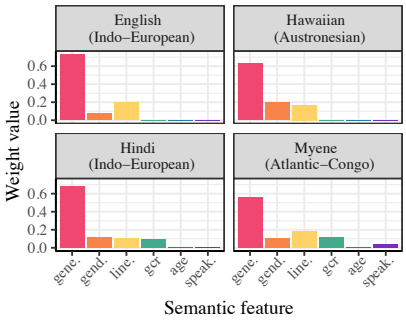


Figure 1. Final weight values for a sample of four languages in our dataset. Generation is reliably weighted highest, but otherwise features vary in the extent to which they cue shared category membership.

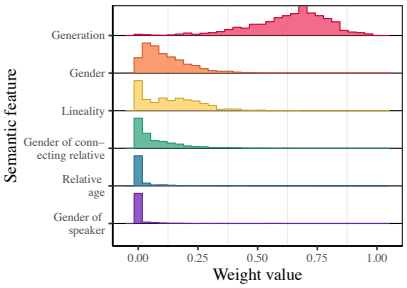


Figure 2. Distribution of final weight values for each semantic feature, ordered by median value. Weights significantly decrease between successive features, suggesting a hierarchical preference for which semantic features cue shared category membership.

tic features are associated with shared wordforms in 1182 languages (data from Kinbank: Passmore et al., 2023) by identifying six semantic features of interest: **Generation:** which generation kin belongs to; **Gender:** gender of kin; **Lineality:** whether kin is in Ego's direct lineage or not; **Gender of Connecting Relative:** gender of individual who connects kin to Ego; **Relative Age:** age of kin relative to their counterpart (e.g. elder vs younger siblings); and **Speaker Gender:** gender of Ego (the person using the term).

We measured topographicity as the correlation between semantic similarity (the weighted proportion of semantic features shared) and wordform similarity (the proportion of each string that is identical, i.e. inverse normalised Levenshtein edit distance: Levenshtein, 1965) for all pairwise combinations of kin terms in a language (e.g. Kirby et al., 2008; Monaghan et al., 2014). We then used gradient descent to maximise this correlation by searching the space of semantic similarity weights, determining the relative contribution of each feature to topographicity. This left us with a metric that specifies how much each feature contributes to structured form-meaning mappings in a particular language; i.e. the extent to which kin who share that feature are referred to by a similar term.

Our results show that languages differ in which semantic features cue shared category membership: Figure 1 shows the optimal weights for each feature for four languages in our dataset. Figure 2 shows that across the full dataset that there is substantial variation between features, but nevertheless demonstrates that this variation is constrained. Accounting for shared ancestry between languages, we found a significant decrease in weight between each successive pair (e.g., Generation is reliably weighted higher than Gender, and Gender is reliably weighted higher than Lineality, etc.), revealing a cross-linguistic hierarchy in the encoding

of semantic features in kin terms. We propose that the relative ordering of features represents an adaptation to efficient communication, a trade-off between how easy a feature is to learn and the amount of information that the feature provides about the referent.

References

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

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *336*(6084), 1049–1054.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *4*(1), 109–128.

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

Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics. Doklady*, *10*, 707-710.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Phil. Trans. Roy. Soc. B. Biological Sciences*, *369*(1651), 20130229.

Passmore, S., Barth, W., Greenhill, S. J., Quinn, K., Sheard, C., Argyriou, P., Birchall, J., Bowern, C., Calladine, J., Deb, A., Diederen, A., Metsäranta, N. P., Araujo, L. H., Schembri, R., Hickey-Hall, J., Honkola, T., Mitchell, A., Poole, L., Rácz, P. M., Roberts, S. G., Ross, R. M., Thomas-Colquhoun, E., Evans, N., & Jordan, F. M. (2023). Kinbank: A global database of kinship terminology. *18*(5), e0283218.

How redundant is language really? Typological and diachronic views on argument disambiguation strategies

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 Linguistics, University of Antwerp, Antwerp, Belgium

Redundant marking of grammatical relations seems to be commonplace across languages, and benefits learning as well as robust information transmission. At the same time, languages also often exhibit strategy trade-offs, suggesting that redundancy may be dispreferred considering communicative efficiency. In this paper, we assess syntagmatic redundancy in light of these competing motivations, focusing on participant role disambiguation in (historical) English and Dutch ditransitives. Drawing on relevant corpus data, we find that redundancy is prevalent, albeit within certain limits.

1. Argument Disambiguation and (Non-)Redundancy

One of the chief tasks that language users face when interpreting (di)transitive utterances is determining who did what to whom. To tackle this task, they commonly draw on four main morphosyntactic strategies, viz. constituent order, nominal marking, verbal agreement, or prepositional marking (Malchukov et al. 2010; Lamers & de Hoop 2005; Lamers & De Swart 2012). However, language comprehenders can typically also rely on semantic-pragmatic biases such as animacy asymmetries between subjects and objects, which provide sufficient information for argument identification in many cases (Bornkessel-Schlesewsky & Schlesewsky 2009; Czypionka et al. 2017; Mahowald et al. 2023).

The systemic availability of multiple morphosyntactic features in addition to semantic-pragmatic cues then raises the question whether – in individual instances – strategies are often used redundantly (syntagmatic redundancy), or whether languages tend towards complementary, trade-off-like strategy use. The latter scenario is suggested by a common assumption in linguistics, namely that grammar is organised in such a way that it facilitates efficient usage (e.g. Gibson et al. 2019; Hahn et al. 2020; 2021). Support for this comes from phenomena such as differential object marking, where e.g. morphosyntactic (prepositional) marking is only applied in contexts where semantic-pragmatic information fails

to disambiguate. On the other hand, languages have been shown to abound with redundant marking of relations within individual utterances even in cases where ambiguity is either clearly resolved by other strategies already, or where there is no clear advantage for language production (Van de Velde 2014; Winter 2014; Levshina 2020, 2021; Tal & Arnon 2022; Tal et al. 2022, 2023). This can be explained by two main benefits of redundancy, viz. robustness against information loss, and learnability.

2. Agent-Recipient Disambiguation in (Historical) English and Dutch

2.1. Set-up of Our Study

In our study, we assess the extent of redundancy in light of the competing motivations of efficiency vs robustness/ learnability, by zooming in the use of multiple morphosyntactic strategies to distinguish agents and recipients in transfer-events with give. We choose this particular case as it lets us investigate formal redundancy without taking into account semanticpragmatic biases, with both arguments being prototypically animate (Newman 1998; Naess 2007; Haspelmath 2015). We focus on the two closely related languages of Dutch and English as they employ the same morphosyntactic strategies but to different degrees, which we predict also affects the extent and specific distribution of redundant marking. Historical English is included as further point of comparison, this time diachronic, since the strategies have seen substantial change over time in this language. To investigate our questions, we use relevant data (N=approx. 1,500) from corpora of English (ICE, Greenbaum & Nelson 1996; GloWbE; Davies 2013), Dutch (Sonar; Oostdijk et al. 2013), and Middle English (PPCME2; Kroch et al. 2000).

2.2. Results

Our results show that redundancy, and specifically double marking, is the default in both Dutch and English. Furthermore, Dutch language users make use of the greater flexibility offered by the system to modulate the degree of redundancy in their sentences in an ad-hoc way, while English grammar exhibits a more consistent level of double marking throughout. In addition, redundancy in Dutch but not English seems to be impacted by processing complexity. Historical English appears to show features of both present day languages, indicating a system in transformation. We interpret these findings as supporting efficiency as a guiding

factor in strategy use, but in a more gradient way, which is impacted, among other things, by language-specific systemic properties.

References

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass, 3*, 19-58. http://doi.org/10.1111/j.1749-818X.2008.00099.x

Czypionka, A., Spalek, K., Wartenburger, I., & Krifka, M. (2017). On the interplay of object animacy and verb type during sentence comprehension in German. *Linguistics, 66*(5), 1383-1433. http://doi.org/10.1515/ling-20170031

Davies, M. (2013). Corpus of Global Web-Based English. https://www.englishcorpora.org/glowbe/

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. https://doi.org/10.1016/j.tics.2019.02.003

Greenbaum, S., & Nelson, G. (1996). The International Corpus of English (ICE) Project. *World Englishes, 15*(1), 3-15. http://doi.org/10.1111/j.1467971X.1996.tb00088.x

Hahn, M., Degen, J., & Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient trade-off of memory and surprisal. *Psychological Review, 128*(4), 726-756. http://doi.org/10.1037/rev0000269

Hahn, M., Jurafsky, D., & Futrell. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences 117*(5), 2347-2353. http://doi.org/10.1073/pnas.1910923117

Haspelmath, M. (2015). Ditransitive constructions. *Annual Review of Linguistics 1*, 19-41. http://doi.org/10.1146/annurev-linguist-030514-125204

Kroch, A., Taylor, A., & Santorini, B. (2000). The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Second edition, release 4. http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCME2-RELEASE4

Lamers, M., & de Hoop, H. (2005). Animacy information in human sentence processing. In H. Christiansen, P. Skadhauge and J. Villadsen (Eds.), *Constraint solving and language processing* (pp. 158-171). Berlin: Springer. http://doi.org/10.1007/11424574_10

Lamers, M., & de Swart, P. (Eds.). (2012). *Case, word order and prominence*. Dordrecht: Springer. http://doi.org/10.1007/978-94-007-1463-2

Levshina, N. (2020). Efficient trade-offs as explanations in functional linguistics: some problems and an alternative proposal. *Revista da Abralin, 19*(3), 50-78. http://doi.org/10.25189/rabralin.v19i3.1728

Levshina, N. (2021). Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related

explanations. *Frontiers in Psychology, 12*, 2791. http://doi.org/10.3389/fpsyg.2021.648200

MacWhinney, B., Malchukov, A., & Moravcsik, E. (Eds.). (2014). *Competing motivations in grammar and usage*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198709848.001.0001

Mahowald, K., Diachek, E., Gibson, E., Fedorenko, E., & Futrell, R. (2023). Grammatical cues to subjecthood are redundant in a majority of simple clauses across languages. *Cognition, 241*, 105543. https://doi.org/10.1016/j.cognition.2023.105543

Malchukov, A., Haspelmath, M., & Comrie, B. (2010). Ditransitive constructions: A typological overview. In A. Malchukov, M. Haspelmath and B. Comrie (Eds.), *Studies in ditransitive constructions* (pp. 1-64). Berlin: De Gruyter Mouton. http://doi.org/10.1515/9783110220377.1

Naess, A. (2007). *Prototypical transitivity*. Amsterdam: Benjamins. http://doi.org/10.1075/tsl.72

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. http://doi.org/10.1075/cagral.3.03new

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. 219247). Heidelberg: Springer. http://doi.org/10.1007/978-3-642-30910-6_13

Tal, S., & Arnon, I. (2022). Redundancy can benefit learning: Evidence from word order and case marking. *Cognition, 224*, 105055. http://doi.org/10.1016/j.cognition.2022.105055

Tal, S., Smith, K., Culbertson, J., Grossman, E., & Arnon, I. (2022). The impact of information structure on the emergence of differential object marking: an experimental study. *Cognitive Science, 46*(3), e13119. http://doi.org/10.1111/cogs.13119

Tal, S., Grossman, E., Rohde, H., & Arnon, I. (2023). Speakers use more redundant references with language learners: Evidence for communicatively efficient referential choice. *Journal of Memory and Language, 128*, 104378. https://doi.org/10.1016/j.jml.2022.104378

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. http://doi.org/10.1515/9783110366273.141

Winter, B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality: Prospects & Overviews. *BioEssays, 36*(10), 960-967. http://doi.org/10.1002/bies.201400028

How and why does the Zipfian distribution differ between spoken and written language

Anna Petrova[1*], Kathleen Rastle[1], Matteo Lisi[1], Shiri Lev-Ari[1]
Corresponding Author: anna.petrova@rhul.ac.uk
[1]Royal Holloway, University of London

Zipf law appears across domains, but its source remains unclear. To reveal its source, we investigated how Zipf's law manifests in spoken versus written language, as the modalities differ in their cognitive demands. Two key finding emerged: spoken language shows a steeper frequency distribution slope, especially under time pressure. Additionally, spoken language exhibits higher semantic similarity between words. This suggests that Zipf's law is motivated by cognitive demands and has implications for prediction and language learning.

The frequency of words in language follows Zipf's law – an inverse power law association between word frequency and rank [1]. This pattern holds across languages [2] but its source is still debated [4]. Our work approaches the question of its source from a new direction by examining differences in the frequency distribution between the spoken and written modalities and the mechanisms underlying these differences.

Written language exhibits higher lexical diversity than spoken language [3], so spoken language's frequency distribution might be steeper. We tested this hypothesis using the Spencer corpus [5], for which participants produced written and spoken content on the same topic in counterbalanced order. A comparison of content from the same individuals, to the same prompt, provides a strong test of whether modality alone can influence the frequency distribution. An analysis of the distribution parameters by MLE showed that the spoken distribution is significantly steeper (spoken slope: $a=1.31$, written slope: $a=1.11$; $p<0.005$; see Fig. 1).

Why is the Zipfian curve steeper in spoken modality? We propose two mechanisms: (1) time pressure limits speakers to semantically accessible words, and (2) the greater cognitive demand that speaking poses [6] leads to increased reliance on high-frequency words.

To test mechanism (1), we calculated the median cosine similarity between the words in each modality in the Spencer corpus using a pre-trained word2vec model as a proxy for semantic accessibility. A linear regression showed that though on the same topic, words in the spoken modality had significantly higher similarity ($F=18.61$, $p<0.005$; see Fig. 2), consistent with speakers drawing from a narrower semantic space under time pressure.

To test mechanism (2) - that cognitive demand leads to higher use of more frequent words, we compared the frequency distributions of 25 conference talks from the Abralin program vs the answers in the Q&A. Thus the two conditions were matched for topic, speaker, and modality and only differed in terms of pressure and preparedness, with Q&A requiring real-time responses under higher cognitive demand. We found a significantly steeper distribution for Q&A than for

talks (talks slope: a=1.12; Q&A slope: a=1.28, p<0.005; see Fig. 3), supporting the cognitive demand hypothesis.

Overall, the studies show that the word frequency distribution is steeper in spoken language, potentially because of the reduced semantic breadth and increased cognitive demand in spoken language production and in spontaneous versus prepared speech. These results contribute to understanding the cognitive mechanisms shaping the shape of the frequency distribution. Furthermore, as language use has been shifting to the written modality, and the shape of the Zipfian distribution influences language learning and use [7], these results suggest that the shift to written communication can affect language learning and use patterns.

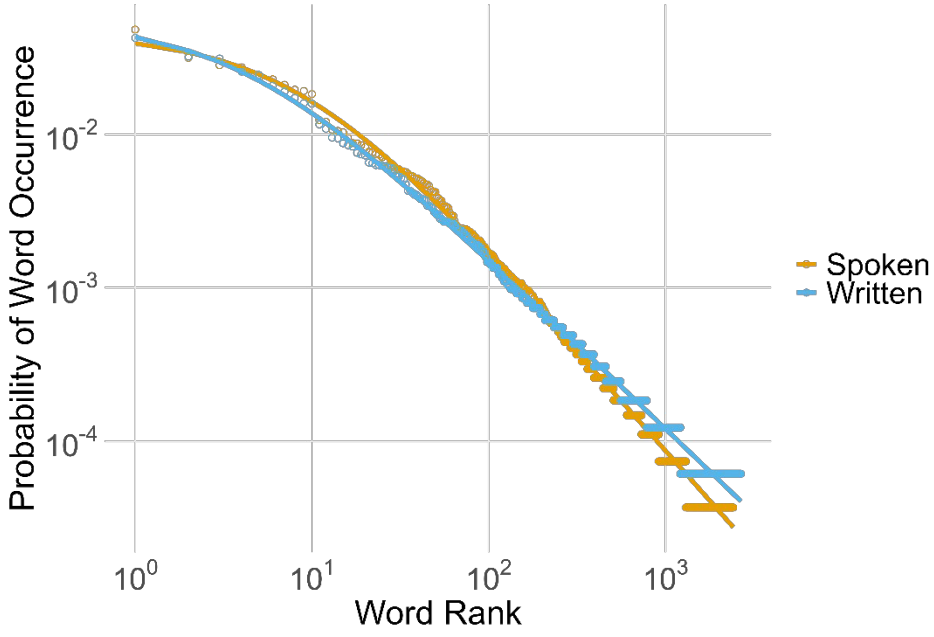


Fig. 1: A comparison of frequency distribution between written and spoken subsets of the Spencer corpus

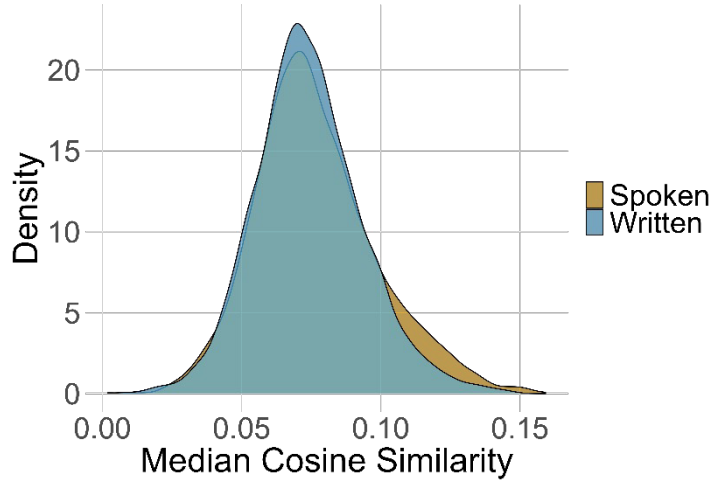


Fig. 2 Median cosine similarity between words in the written and spoken subsets of the Spencer corpus

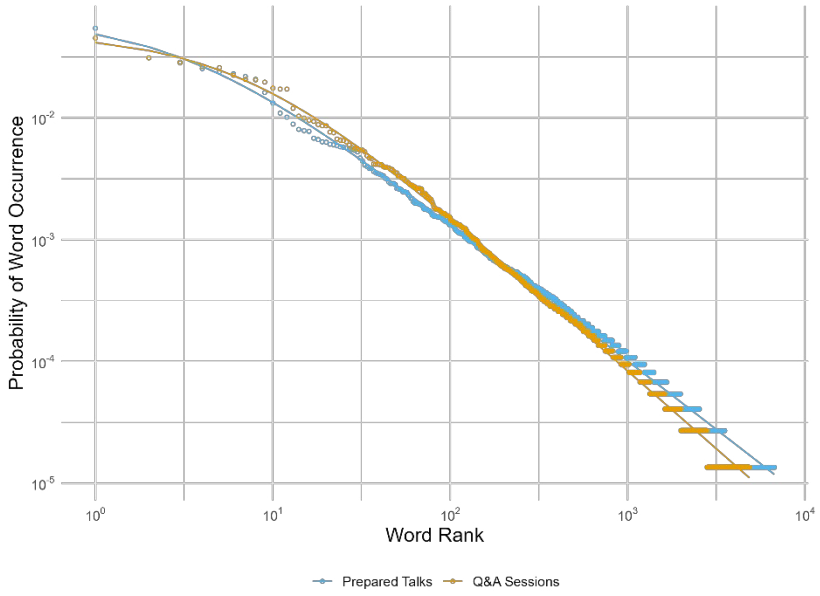


Fig. 3 A comparison of frequency distribution between conference talks and Q&A answers

Acknowledgements
This research was supported by Leverhulme Trust grant R11495.

References

[1] Zipf, G. K. (1949). The psychobiology of language: An Introduction to dynamic philology.
[2] 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*.
[3] Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied linguistics*, *31*(2), 236-259.
[4] Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, *21*, 1112-1130.
[5] Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy*, *5*(1), 1-43.
[6] Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.
[7] Lavi-Rotbain, O., & Arnon, I. (2022). The learnability consequences of Zipfian distributions in language. *Cognition*, *223*, 105038.

How precisely does human eye colour characterization follow colour naming convention: An experimental study

Vojtech Fiala[*1,2], Juan Olvido Perea-García[1,3], Anna Szala[1], Theresa Matzinger[4], Karel Kleisner[2], Jindřich Brejcha[2], Przemysław Żywiczyński[1], Sławomir Wacewicz[1]

[*]Corresponding Author: vfiala@umk.pl

[1]Centre for Language Evolution Studies, Nicolaus Copernicus University, Toruń, Polska

[2]Department of Philosophy and History of Science, Charles University, Prague, Czechia

[3]Institute of Biomedical and Health Research, University of Las Palmas de Gran Canaria, Las Palmas, Spain

[4]Department of English and American Studies, University of Vienna, Vienna, Austria

Eye (i.e. iris) colour is a salient personal trait for which language may assign naming patterns different than for other object categories. To test this possibility, we developed a survey in which participants specify the colour of the stimuli irises in a questionnaire and select the colour itself in a colour chart. We investigate to what extent labelling eye colour is top-down (driven by linguistically established eye colour categories) vs bottom-up (driven by the stimulus' low-level psychophysiological properties).

1. Introduction

Compared to other mammals, colour perception is particularly well-developed among Old World primates, including humans (Vorobyev, 2004; Jacobs, 1993). Even though our biology constrains colour perception (Skelton et al., 2017), there is substantial variation in colour categorisation across individuals and populations (Gibson et al., 2017). Beyond individual differences that may cause variations in chromatic perception of the same colour (e.g. Bosten, 2022; Emery & Webster, 2019; Paramei et al. 2004), research suggests that both culture and language systematically influence how speakers categorize and cognitively process perceived colours (Özgen et al., 2004, cf. Kay & Reiger, 2006). Similarly, some culture-related factors (local conditions, lack of knowledge of foreign facial morphology) have been claimed to lower agreement in facial characterisation during first impression formation (Pavlovič et al., 2021; Sorokowski et al., 2013). Together, lack of knowledge, different adaptations, and variability in naming conventions may lead to cross-population variations in the perception and verbal characterisation of certain facial features.

Eye colour naming patterns may present such an example: The cross-cultural agreement can be lowered both by the lack of knowledge of different eye morphologies and colour categorisation differences. However, it may also present a special case of a cross-culturally stable naming convention. When Fiala et al. (2024) asked participants from distant countries to distinguish eye colour into

simple categories (“blue”, “brown”, “other”), they responded quite congruently, and eye colour variance measured in CIELab predicted assigned eye colours similarly across cultures.

Even if certain iris colours were absent in a population due to limited iris colour variance (e.g., subequatorial Africa, South-East Asia) and related expressions were historically uncommon, terms like 'blue/brown' and 'light/dark eyes' may have been adopted through exposure to globalised culture. They may also be understood via nonhuman animal species’ eye colour, which varies largely in (sub)tropical mammals (Tabin & Chiasson, 2024; Perea-García et al. 2022). The highly salient and cross-culturally appreciated eye morphology (Wiącek, 2015) may then facilitate the creation of naming patterns that diverge from the naming pattern for other surface colours and that is partly, or completely invariant between individuals and cultures.

2. The current study

We asked the following questions: **(i)** Is it just the “umbrella concepts” like “blue vs. brown” or “dark vs. light” that are shared between individuals and, potentially, cross-culturally? The agreement may be limited to simple iris colour categories, but it can remain even for detailed categorisation. **(ii)** When participants are asked to pick the iris colour from a chart, what is the distribution of, and cross-cultural agreement in, the colours selected?

We identified a list of common human eye colour names being used to describe iris colour variance in English scientific literature (Mackey et al., 2011), obtained RGB values of these colours, or their best-fitting analogues according to ISCC-NBS system (Judd & Kelly, 1939), converted the values to perceptually uniform CIELab L*a*b colour space (McLaren, 1976), and created a corresponding 2D colour chart with a range of the candidate iris-related colours. Using faces of a population with highly variable iris colour as stimuli (N = 195), we will ask participants to choose the dominant iris colour before selecting it from the chart.

This way, we intend to test whether participants agree on the iris colour naming, on the colour selected from the chart and whether the level of congruence and variance depends on the country of participants’ origin and contemporary residence. We will code the names chosen by participants into appropriate categories, and measure variance in naming and variance in selection from the plot (the latter using Manhattan distance). We will analyse the data with Poisson generalised models. We expect that participants will generally agree on the colour they pick; however, there will be a larger variance and distance in raters originating outside Europe. The survey is currently being prepared and will be distributed in Europe, South-East Asia, and (via Prolific) through English-

speaking non-European regions, in English and in languages with similar colour grids (at this stage, we do not plan to consider differences due to a different colour grid in different languages). Results will be available at the time of the conference.

3. Supplementary Materials

https://osf.io/pwajh/?view_only=0b2d7df5dbda4bc6a88dd151eadb1c4c

Acknowledgements

Karel Kleisner and Jindřich Brejcha were supported by the Czech Science Foundation (project no. 24-11720S), Sławomir Wacewicz was supported by the National Science Centre of Poland, grant 2023/49/B/HS6/02343. Przemysław Żywiczyński was supported by the National Science Centre of Poland under the agreement UMO 2023/49/B/HS2/02222.

References

Bosten, J. M. (2022). Do you see what I see? Diversity in human color perception. *Annual Review of Vision Science, 8*(1), 101–133.

Emery, K. J., & Webster, M. A. (2019). Individual differences and their implications for color perception. *Current Opinion in Behavioral Sciences, 30*, 28–33.

Fiala, V., Kleisner, K., Żywiczyński, P., Tureček, P., Wacewicz, S., Štěrbová, Z., & Matzinger, T. (2024, September). Different culture, similar view: Congruence in eye colour perception in WEIRD and non-WEIRD samples. *Paper presented at the Polish Society for Human Evolution Conference*, Katowice, Poland.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., … & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences, 114*(40), 10785–10790.

Jacobs, G. H. (1993). The distribution and nature of colour vision among the mammals. *Biological Reviews of the Cambridge Philosophical Society, 68*(3), 413–471.

Judd, D. B., & Kelly, K. L. (1939). Method of designating colors. *Journal of Research of the National Bureau of Standards, 23,* 355-385

Kay, P., & Regier, T. (2006). Language, thought, and color: Recent developments. *Trends in Cognitive Sciences, 10*(2), 51–54.

Mackey, D. A., Wilkinson, C. H., Kearns, L. S., & Hewitt, A. W. (2011). Classification of iris colour: review and refinement of a classification schema. *Clinical & experimental ophthalmology*, *39*(5), 462-471.

McLaren, K. (1976). XIII—The development of the CIE 1976 (L* a* b*) uniform colour space and colour-difference formula. *Journal of the Society of Dyers and Colourists*, *92*(9), 338-341.

Özgen, E. (2004). Language, learning, and color perception. *Current Directions in Psychological Science, 13*(3), 95–98.

Paramei, G. V., Bimler, D. L., & Mislavskaia, N. O. (2004). Colour perception in twins: Individual variation beyond common genetic inheritance. *Clinical and Experimental Optometry, 87*(4–5), 305–312.

Pavlovič, O., Fiala, V., & Kleisner, K. (2021). Environmental convergence in facial preferences: A cross-group comparison of Asian Vietnamese, Czech Vietnamese, and Czechs. *Scientific Reports, 11*, 550.

Perea-García, J. O., Ramarajan, K., & Kret, M. E. (2022). Ecological factors are likely drivers of eye shape and colour pattern variations across anthropoid primates. *Scientific Reports, 12*, 17240.

Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M., & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences, 114*(21), 5545–5550.

Sorokowski, P., Kościński, K., & Sorokowska, A. (2013). Is beauty in the eye of the beholder but ugliness culturally universal? Facial preferences of Polish and Yali (Papua) people. *Evolutionary Psychology, 11*(4), 907–925.

Tabin, J. A., & Chiasson, K. A. (2024). Evolutionary insights into *Felidae* iris color through ancestral state reconstruction. *iScience, 27*(10)

Vorobyev, M. (2004). Ecology and evolution of primate colour vision. *Clinical and Experimental Optometry, 87*(4–5), 230–238.

Wiącek, E. (2015). Old symbol in contemporary culture: Turkish evil eye bead: From ancient amulets to commercial gifts. In *Symbols of contemporary culture* (pp. 15–22). Jagiellonian University in Kraków, Kraków, Poland.

Iconicity and compositionality can co-develop in communication across modalities

Danielle Naegeli[*1,2], Oxana Grosseck[2], Kotryna Motiekatytė[2], Hans Rutger Bosker[3], Gerardo Ortega[4], Marcus Perlman[4], David Peeters[1,5], Limor Raviv[2,3]

[*]Corresponding Author: d.s.u.naegeli@tilburguniversity.edu

[1]Department of Communication and Cognition, Tilburg University, Tilburg, NL

[2]LEADS group, Max Planck Institute for Psycholinguistics, Nijmegen, NL

[3]Donders Centre for Cognition, Radboud University, Nijmegen, NL

[4]Department of English Language and Linguistics, University of Birmingham, Birmingham, UK

[5]Multimodal Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, NL

Previous work suggested that gestural and vocal communication show different degrees of iconicity, and that high iconicity may hinder the development of compositional structure. Here, we tested the evolution of iconicity and compositionality across modalities using a VR communication game. Results show that gesture-only pairs were more iconic and communicatively successful than vocalization-only pairs, yet pairs in both conditions developed compositional languages, often reusing iconic elements. Our findings support the idea that iconicity facilitates the emergence of structure across modalities.

Iconicity can help bootstrap emerging communication (e.g., Fay et al., 2013; Perlman et al., 2015). It is widely present in gestures, and recent work shows that it is also more prevalent in the vocal modality than previously assumed (e.g., Dingemanse et al., 2015; Perlman, 2017). At the same time, iconicity in acoustic signals seems to hinder the development of compositional structure (Verhoef et al., 2016), which is considered central to language evolution (Kirby et al., 2008, Motamedi et al., 2019). Crucially, only few studies have directly compared novel signal creation in the vocal and gestural modalities (Fay et al., 2013, 2014, 2022; Lister et al., 2021; Macuch-Silva et al., 2020; Motiekatytė et al., 2024), and no study to date has examined how compositionality evolves over time across modalities and how it relates to iconicity in manual vs. vocal communication.

To test the (co-)evolution of iconicity and compositionality across modalities, we conducted a dyadic communication game in an immersive Virtual Reality CAVE environment (Cruz-Neira et al., 1992). Pairs of participants interacted

face-to-face using either gesture or non-linguistic vocalizations to refer to novel creatures appearing around them in a virtual forest. The stimuli consisted of 32 fantasy creatures varying on four semantic features: shape (4 different creatures), size (small vs. big), movement (walk vs. jump), and speed (fast vs. slow). We used PRAAT and ELAN to annotate the emerging communication systems of 6 vocalization-only and 6 gesture-only pairs, analyzing signals for their duration, number of syllables/gestures, speech/gesture rate (i.e., signal duration divided by number of syllables/gestures), iconicity (measured as guessing accuracy by naive participants; Grosseck et al., 2024), and compositionality (measured as pair-wise correlations between meanings and signal annotations).

Preliminary results showed that gesture-only pairs communicated more successfully and created more iconic signals than vocal-only pairs (replicating Fay et al., 2014; Macuch-Silva et al., 2020), yet accuracy and iconicity were also significantly above chance in vocal-only pairs (Grosseck et al., 2024; Motiekatytė et al., 2024). Importantly, we found that compositionality emerged to different degrees in both modalities and increased over time, with different pairs systematically reusing and combining iconic gestures or vocalizations to denote different semantic features. For example, creatures' speed was typically encoded in the production rate of signals (i.e., faster movements expressed by faster vocalizations and gestures). Notably, iconicity remained stable over time, and emerging compositional languages often relied on iconicity (see Figure 1). These results show that iconicity and compositionality can emerge through communication alone (extending Motamedi et al., 2019). Furthermore, our findings support the idea that iconicity and compositionality can co-exist (Little 2017a, 2017b).

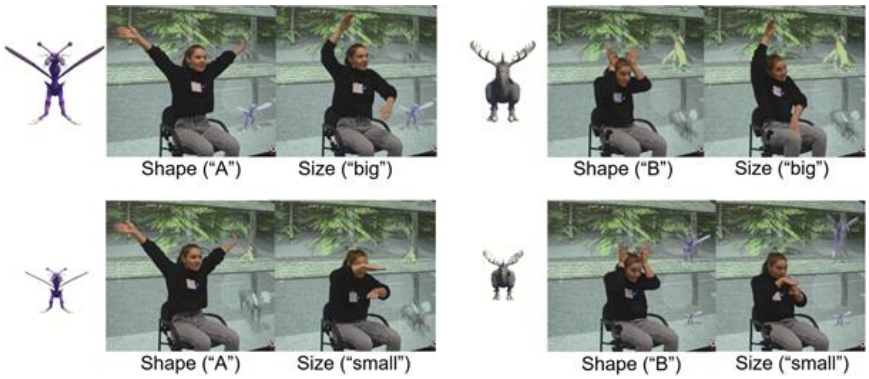


Figure 1: Example of a compositional system that reuses iconic gestures for size and shape.

References

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM*, *35*(6), 64-72.

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.

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive Science*, *37*(7), 1356-1367.

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014). Creating a communication system from scratch: gesture beats vocalization hands down. *Frontiers in Psychology*, *5*, 354.

Fay, N., Walker, B., Ellison, T. M., Blundell, Z., De Kleine, N., Garde, M., ... & Goldin-Meadow, S. (2022). Gesture is the primary modality for language creation. *Proceedings of the Royal Society B*, *289*(1970), 20220066.

Grosseck, O., Perlman, M., Ortega, G., & Raviv, L. (2024). 'The iconic affordances of gesture and vocalization in emerging languages in the lab. *The Evolution of Language: Proceedings of the 15th International Conference on the Evolution of Language (EVOLANG XV)*, 223-225.

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.

Little, H., Eryılmaz, K., & de Boer, B. (2017a). Signal dimensionality and the emergence of combinatorial structure. *Cognition*, *168*, 1–15. https://doi.org/10.1016/j.cognition.2017.06.011

Little, H., Perlman, M., & Eryılmaz, K. (2017b). Repeated Interactions Can Lead to More Iconic Signals. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *39*(0). https://escholarship.org/uc/item/1pk2q9n4

Lister, C. J., Burtenshaw, T., Walker, B., Ohan, J. L., & Fay, N. (2021). A Cross-Sectional Test of Sign Creation by Children in the Gesture and Vocal Modalities. *Child Development*, *92*(6), 2395-2412.

Macuch Silva, V., Holler, J., Özyürek, 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.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign language in the lab: From improvised gesture to systematic sign. *Cognition*, *192*, 103964.

Motiekatytė, K., Grosseck, O., Wolf, L., Bosker, H. R., Peeters, D., Perlman, M., Ortega, G., & Raviv, L. (2024). 'The iconic affordances of gesture and vocalization in emerging languages in the lab. *The Evolution of Language: Proceedings of the 15th International Conference on the Evolution of Language (EVOLANG XV)*, 387-389.

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity Can Ground the Creation of Vocal Symbols. *Royal Society Open Science*, *2*(8): 150152.

Perlman, M. (2017). Debunking two myths against vocal origins of language: language is iconic and multimodal to the core. *Interaction Studies*, *18*(3), 376- 401.

Verhoef, T., Kirby, S., & De Boer, B. (2016). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science*, *40*(8), 1969-1994

Investigating intentionality in elephant gestural communication

Vesta Eleuteri[1*], Lucy Bates[2], Yvonne Nyaradzo Masarira[3], Joshua M. Plotnik[4,5], Catherine Hobaiter[6], Angela Stöger-Horwath[1,7]

*Corresponding author: vesta.eleuteri@univie.ac.at

[1] Department of Behavioral & Cognitive Biology, University of Vienna, Vienna, Austria

[2] Department of Psychology, University of Portsmouth, Portsmouth, United Kingdom

[3] Elephant CREW, Jafuta Reserve, Zimbabwe

[4] Department of Psychology, Hunter College, City University of New York, New York, NY, USA

[5] Department of Psychology, Graduate Center, City University of New York, New York, NY, USA

[6] School of Psychology and Neuroscience, University of St Andrews, St Andrews, United Kingdom

[7] Acoustic Research Institute, Austrian Academy of Sciences, Vienna, Austria

Goal-directed intentionality is a defining feature of language and represents the ability to communicate goals in mind to others. To establish goal-directed intentionality (hereafter intentionality), key criteria need to be met: audience directedness, persistence, or elaboration. While intentionality has been shown in a few specific gesture types in non-primates, great apes are known to use large gestural repertoires intentionally. But is communicating intentionally with many gesture types unique to primates? Here we show that semi-captive elephants use many gestures intentionally to ask humans for food.

1. Introduction

Humans use language *intentionally* to communicate cognitive goals (i.e., meanings) eliciting specific behavioural reactions in recipients (Dennett, 1983; Grice, 1957). To establish intentionality, the following criteria need to be met (Tomasello et al., 1985; Townsend et al., 2017): a) audience directedness – signallers use gestures in the presence of a perceiving audience; b) persistence - signallers persist gesturing when recipients do not meet their goal; or c) elaboration - signallers change gesture types when previous ones fail at achieving their goal. Apes use large gestural repertoires intentionally by meeting these

criteria (Byrne et al., 2017). Beyond primates, intentional gesturing is reported in a few species, for example coral reef fishes, Arabian ballers, and ravens, but only in a few gesture types (Ben Mocha & Burkart, 2021; Pika & Bugnyar, 2011; Vail et al., 2013). Whether non-primates communicate intentionally across many gesture types remains unknown. Here, we built on Leavens and colleagues' experimental design (2005) to test whether semi-captive elephants gesture intentionally to a human experimenter to acquire food.

2. Methods

Elephants were 17 semi-captive adult African savannah elephants (8 males and 9 females). We presented elephants with unreachable desired and non-desired items. The desired item was a tray containing 6 apples, while the non-desired item was an empty tray. Each elephant was presented once with three experimental trials corresponding to: 1) **Goal met condition**: The experimenter delivered all baited apples from the tray. 2) **Goal not met condition**: The experimenter delivered the empty tray. 3) **Goal partially met condition**: The experimenter delivered one of the baited apples from the apple tray. The trials consisted in three experimental phases. **Pre-delivery phase:** the two items were positioned in front of the elephant by a helper who then left. The experimenter entered and positioned themselves in-between the trays and faced the elephant for 40 seconds. After the time elapsed, the experimenter delivered one of the items according to condition (e.g., delivered all apples in Goal met condition). **Post-delivery phase**: after delivery, the experimenter went back in-between the trays and faced the elephant for another 40 seconds. **No-experimenter phase:** After the time elapsed the experimenter left, and the elephant was left alone for another 40 seconds. The trial then ended. Trials were pseudorandomised and counterbalanced across elephants.

3. Results

Elephants used a total of 313 gestures of 38 gesture types across 51 experimental trials. Elephants gestured only in the presence of the experimenter (audience directedness). Elephants increased their use of gestures (i.e., persisted more) from pre-delivery to post-delivery in the Goal partially met condition than in the Goal met condition (persistence). But, interestingly, during post-delivery, elephants decreased gesturing in both the Goal not met and Goal met conditions, and no difference in gesture persistence was found between these conditions. However,

during post-delivery, elephants used novel gesture types more often in the Goal not met condition compared to the Goal met condition (elaboration).

4. Discussion

Apes are known to gesture intentionally across large gestural repertoires and evidence of intentionality in non-primates is restricted to one or two gesture types selected for specific purposes (Ben Mocha & Burkart, 2021; Pika & Bugnyar, 2011; Vail et al., 2013). Here we show that semi-captive elephants use 38 different gesture types only in the presence of an attentive human audience, that they persist gesturing more when their goal is partially met compared to fully met, while elaborate by using new gesture types after previous ones completely failed to meet their goal. Future studies should investigate intentional gesturing between elephants to further our understanding of the evolution of this capacity across socially complex distantly related species.

Acknowledgements

We thank the elephant carers and all the staff at elephant CREW and Wild Horizons for their assistance in the field. We thank Anna Letrari for assistance in data collection. This research was funded by the Austrian Science Fund (FWF) [W1262-B29 [10.55776/W1262]] and the European Union's 8th Framework Programme, Horizon 2020 [802719].

References

Bates, L. A. (2020). Cognitive Abilities in Elephants. In J. H. Barkow, L. Workman, & W. Reader (Eds.), *The Cambridge Handbook of Evolutionary Perspectives on Human Behavior* (pp. 14–22). Cambridge University Press. https://doi.org/10.1017/9781108131797.003

Ben Mocha, Y., & Burkart, J. M. (2021). Intentional communication: Solving methodological issues to assigning first-order intentional signalling. *Biological Reviews*, *96*(3), 903–921. https://doi.org/10.1111/brv.12685

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

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *The Behavioral and Brain Sciences*, *6*, 343–390.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388. https://doi.org/10.2307/2182440

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as Measured in the Persistence and Elaboration of Communication by Chimpanzees (Pan troglodytes). *Child Development*, *76*(1), 291–306. https://doi.org/10.1111/j.1467-8624.2005.00845.x

Moss, C. J., Croze, H., & Lee, P. C. (2011). *The Amboseli Elephants: A Long-Term Perspective on a Long-Lived Mammal*. University of Chicago Press.

Pika, S., & Bugnyar, T. (2011). The use of referential gestures in ravens (Corvus corax) in the wild. *Nature Communications*, *2*(1). https://doi.org/10.1038/ncomms1567

Tomasello, M., George, B. L., Kruger, A. C., Jeffrey, M., Farrar, & Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution*, *14*(2), 175–186.

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., … Manser, M. B. (2017). Exorcising Grice's ghost: An empirical approach to studying intentional communication in animals: Intentional communication in animals. *Biological Reviews*, *92*(3), 1427–1433. https://doi.org/10.1111/brv.12289

Vail, A. L., Manica, A., & Bshary, R. (2013). Referential gestures in fish collaborative hunting. *Nature Communications*, *4*(1). https://doi.org/10.1038/ncomms2781

Language endangerment and biodiversity threat both correlate with duration of European colonial occupation

Bernd Lenzner[*1], Andreas Baumann[2,3], Sietze Norder[4], Franz Essl[1], and Hannes Fellner[3,5,6]

[*]Corresponding Author: bernd.lenzner@univie.ac.at
[1]Department of Botany and Biodiversity Research, University of Vienna, Austria
[2]Department of German Studies, University of Vienna, Austria[3]Other Research
[3]Network Data Science, University of Vienna, Austria
[4]Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, The Netherlands
[5]Department of Linguistics, University of Vienna, Austria
[6]Department of European and Comparative Literature and Language Studies

Biodiversity and language diversity have been suggested to be impacted by anthropogenetic factors. In this paper, we investigate to what extent the duration under European colonial rule is associated with biodiversity threat and language diversity threat by deriving the IUCN Red List Indicator (RLI) on the country level for both domains. We show that, on a global scale, duration of occupation correlates with biodiversity and language threat.

Both biodiversity and linguistic diversity are under threat. About one million species and roughly 3000 languages show a risk of extinction (Johnson et al. 2017; Bromham et al. 2022). Some of the causes for this are anthropogenetic, such as overexploitation and increased land usage, both of which have been argued to have negatively affected biota and indigenous languages (Sutherland 2003; Skutnabb-Kangas & Harmon 2017). A long-term event that was evidently associated with processes like this is that of European colonial expansion (Nettle & Romaine 2000; Mufwene 2004). In our paper, we directly compare biodiversity threat and threat of linguistic diversity against each other on a global scale, and we examine to what extent the duration of European colonial occupation is associated with threat in both domains. Our results indicate that biodiversity and language diversity are conditioned by different factors, but that European colonialism displays a negative association with both of them.

To make biodiversity threat and threat of linguistic diversity comparable, we employ the IUCN Red List Indicator (RLI), which is an established framework in

biodiversity research. RLI maps average threat levels for a particular region onto a normalized scale so that 0 and 1 indicate that all species in that region are extinct and stable, respectively. We use IUCN Red List assessments for four classes of species (mammals, birds, amphibians, and reptiles), and EGIDS threat levels for languages as provided by Ethnologue. The scales were homogenized to yield the same number of threat levels (five) so that they could be mapped onto each other. We then computed $RLI_{biological}$ and $RLI_{linguistic}$ for each country as

$$RLI_d = 1 - \sum_i \frac{T_i}{T_{max}}$$

by summing over the (normalized) threat levels T_i of all species/languages in that country (Loh & Harmon 2014) for both domains separately (where d stands for the two domains). In addition, we computed a composite index $RLI_{biocultural}$ as the mean of $RLI_{biological}$ and $RLI_{linguistic}$, which can function as a proxy of biocultural diversity on a more general level.

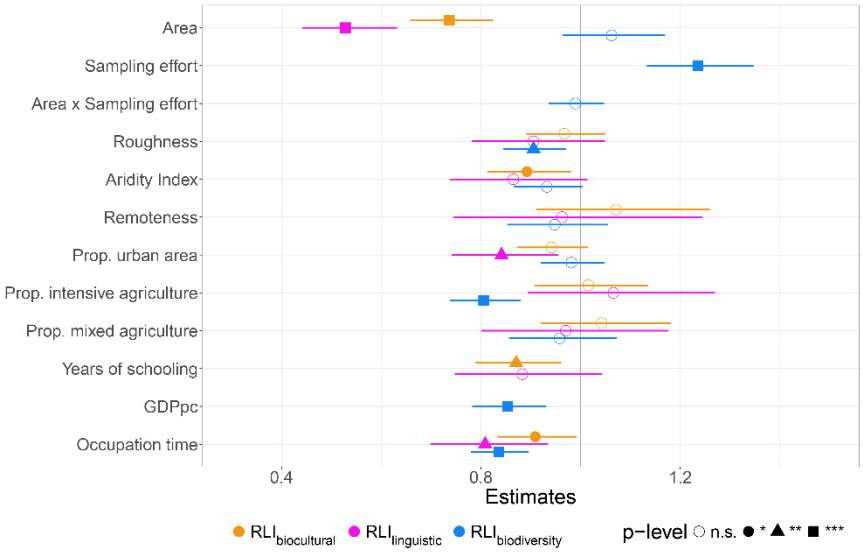


Figure 1. Coefficients together with 95% confidence intervals for the three beta-regression models. Response variables: $RLI_{biocultural}$ (orange), $RLI_{linguistic}$ (purple), $RLI_{biological}$ (blue).

In the first part of our analysis, we measure the association between $RLI_{biological}$ and $RLI_{linguistic}$, and identify hotspots of biological and linguistic diversity threat, controlling for country size (area). After that, we fit three beta regression models, one for each RLI measure as response variable and occupation time as predictor. Occupation time was derived from the COLDAT database (Becker 2019), which provides beginning and end dates of European occupation for each country. In addition, we use a set of other covariates that have been suggested to be relevant

to threat of biodiversity and language diversity. All predictor variables were z-transformed. We find that, although there are differential effects of the other covariates, occupation time is negatively associated with all three RLI measures (Figure 1; effect sizes of occupation time: $\beta = -0.09^{*}$ for $\text{RLI}_{\text{biocultural}}$; $\beta = -0.18^{*}$ for $\text{RLI}_{\text{biological}}$; $\beta = -0.21^{*}$ for $\text{RLI}_{\text{linguistic}}$); longer colonization duration corresponds to higher threat across domains. Our conjecture is that European colonization has reshaped both the socio-cultural and biological environment in such a way that many languages and species are subject to increased chances of extinction. However, differential sets of mechanism apply in both domains.

References

Becker, B. (2019). Introducing COLDAT: the colonial dates dataset.

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., ... & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. Nature ecology & evolution, 6(2), 163-173.

Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., & Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. Science, 356(6335), 270-275.

Loh, J., & Harmon, D. (2014). Biocultural diversity: Threatened species, endangered languages.

Mufwene, S. S. (2004). The ecology of language evolution. Cambridge University Press.

Nettle, D., & Romaine, S. (2000). Vanishing voices: The extinction of the world's languages. Oxford: Oxford University Press.

Skutnabb-Kangas, T., & Harmon, D. (2017). Biological diversity and language diversity: parallels and differences. In The Routledge handbook of ecolinguistics (pp. 11-25). Routledge.

Least but not lost. Finding the requirements for gradual reduction using an agent-based model of least effort

Anthe Sevenants*[1], Dirk Speelman[1], Freek Van de Velde[1], and Dirk Pijpops[2]

*Corresponding Author: anthe.sevenants@kuleuven.be
[1]Department of Linguistics, KU Leuven, Leuven, Belgium
[2]Department of Linguistics, University of Antwerp, Antwerp, Belgium

Humans tend to use the least effort possible when communicating, which leads to the reduction of frequently used phrases over time ("reducing effect"). While reduction benefits speakers, it must operate without causing communicative chaos. We built an agent-based model to simulate this process, and found that communicative stability is only guaranteed when speakers estimate their ability to be understood as hearers, echoing Hockett and Hockett (1960). This requirement highlights how defining principles of human communication also define its further evolution.

Humans are notoriously lazy. Even though communication systems are known to evolve spontaneously out of a common need between speakers and hearers (Oliphant, 1997), the process of linguistic selection (Steels, 2012) states that speakers will still select utterances that reach common communicative need with as little effort as possible. That principle of least effort (Zipf, 1949) translates into speakers using the least effort possible in their pronunciation, as long as the hearer still comprehends their message ("communicative success").

One instantiation of this principle is the reducing effect (Bybee, 2006). This theory states that words with high frequency are eligible for a fast process of phonetic erosion, causing distinctive sound information to be removed from a construction. An example of this is the highly frequent English *dunno*, which eroded from the sequence *I don't know*. Reduction can also happen to lower frequency forms, but to a smaller extent and at a slower pace.

The reduction process is beneficial to speakers, as it allows them to consume less energy in reaching their communicative goals. However, the reducing effect also shows how linguistic selection affects language as a whole, as reduced forms can also become conventionalised within the linguistic system itself. Through self-organisation (Steels, 2012), the system can reorganise itself to express the same meanings using less complex representations. The self-regulatory aspect of language is a reason why language is typically characterised as a complex-adaptive system (Kretzschmar, 2015), which means language lacks a central authority, and arises out of local interactions which shape the entire system.

The reducing effect raises interesting questions about the necessary requirements for speakers to maintain communicative success throughout the reduction process. If the hearer misinterprets the form spoken by the speaker, this can cascade into communicative chaos. Since language has no central control, this chaos avoidance has to arise out of simple rules implicit in local communication.

To find the minimal requirements for successful reduction, we built an agent-based model in Python with MESA (Kazil, Masad, & Crooks, 2020). In our model, we let a society of virtual speakers "talk" to each other, and have the words they use follow the distribution typical of natural language: a few highly frequent words with a long tail of more infrequent words (Zipf, 1965). Each language user has a memory of words with multiple exemplars (Pierrehumbert, 2001), represented by vector representations. Such vector representations are popular in the field of machine learning, to represent both meaning (Mikolov, Chen, Corrado, & Dean, 2013) and acoustic information (Baevski, Zhou, Mohamed, & Auli, 2020). When a language user "speaks" to a hearer, a random exemplar is selected from the speaker's repertoire. That vector is compared to the vector repository of the hearer. With the cosine distance metric, a vector neighbourhood from the hearer's repertoire is selected, and its associated words are tallied. The word most represented in the neighbourhood is the word "heard" by the hearer. Note that the hearer does not know which word the speaker intended, so mishearings are possible. Of course, communication is more complicated in real life, but focusing on the core of the problem is standard practice in any type of modelling (Smaldino, 2023).

To simulate the loss of acoustic information, we allow for a speaker to lower the values in an exemplar's vector at speech time. With this operation, the information that was initially represented in that vector is expressed using less energy. As words are used more, they have the chance to gradually erode into more efficient representations.

Through our simulation, we found that five requirements are imperative for reduction to occur in the way described in empirical studies:

1. Language users should spontaneously reduce forms;
2. There should be shared vocabulary;
3. Vocabulary frequency should follow Zipf's law;
4. Language users should be able to remember multiple forms heard of the same word;
5. Speakers should check whether they would understand a reduced form themselves before speaking.

The first requirement is necessary to have reduction at all. While the next three requirements come naturally from usage-based linguistics, we nonetheless

disabled them iteratively to check whether they were really necessary to give rise to the reducing effect. (2) If agents use wholly different acoustic forms to refer to the same concepts, communication is impossible; (3) if words do not follow a Zipfian distribution, reduction is equally strong regardless of frequency, which is not what we see in corpus data; (4) if language users only know a single exemplar for each word, the representation of very frequent words shifts too fast, and becomes too varied among agents to still lead to successful communication.

The final requirement is the most interesting theoretically. If the requirement is left out, the language system evolves into a state in which most reduced words are consistently confused with the most frequent word. This is due to the available phonetic space becoming more restricted as more words become reduced, with the most frequent words taking up the most exemplar space due to their sheer frequency. As less frequent words creep into the busy phonetic space, their representations become outnumbered by more frequent exemplars in the neighbourhood. This leads to mishearings and mass communicative confusion.

This communicative confusion can be avoided by having speakers check first whether they would correctly categorise the reduced form they are about to speak themselves. If not, the reduction process is reversed. This simple step avoids most wrong classifications and is roughly related to interchangeability as defined by Hockett and Hockett (1960). This requirement highlights how defining principles of human communication also define its further evolution. More generally, the interaction of our five simple requirements proves that even complex behaviour can be brought back to a few simple defining features.

Acknowledgements

We thank the audience of the "Bridging the Gap" workshop (Zürich, March 2025) for their comments on and suggestions with a previous version of our model.

References

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020, October). *wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations.* arXiv.

Bybee, J. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, *82*(4), 711–733.

Hockett, C. F., & Hockett, C. D. (1960). The Origin of Speech. *Scientific American*, *203*(3), 88–97.

Kazil, J., Masad, D., & Crooks, A. (2020). Utilizing python for agent-based modeling: The mesa framework. In R. Thomson, H. Bisgin, C. Dancy, A. Hyder, & M. Hussain (Eds.), *Social, cultural, and behavioral modeling* (pp. 308–317). Cham: Springer International Publishing.

Kretzschmar, W. A. (2015). *Language and Complex Systems.* Cambridge: Cambridge University Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, September). *Efficient Estimation of Word Representations in Vector Space.* arXiv.

Oliphant, M. W. (1997). *Formal approaches to innate and learned communication: Laying the foundation for language.* Unpublished doctoral dissertation, University of California - San Diego, San Diego.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (Vol. 45, pp. 137–158). Amsterdam: John Benjamins Publishing Company.

Smaldino, P. (2023). *Modeling Social Behavior: Mathematical and Agent-Based Models of Social Dynamics and Cultural Evolution.* Princeton University Press.

Steels, L. (2012). Interactions between cultural, social and biological explanations for language evolution. *Physics of Life Reviews*, *9*(1), 5–8.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Oxford, England: Addison-Wesley Press. (Pages: xi, 573)

Zipf, G. K. (1965). *The Psycho-Biology of Language.* Cambridge, Massachusetts: MIT Press.

Linguistic diversity as a barrier to disease spread

Sihan Chen*[1] and Antonio Benítez-Burraco[2]

*Corresponding Author: sihanc@mit.edu
[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Department of Spanish, Linguistics, and Theory of Literature, Univerity of Seville, Seville, Spain

We test the hypothesis in the literature that linguistic diversity hinders disease spread: people speaking different languages contact less. We first analyze Covid-19 transmission data in countries differing in linguistic diversity, controlling for factors such as development level, geography, and climate. We then simulate the disease transmission in societies varying in size, degree of contact, and network density - three sociopolitical factors impacting language diversity. Both studies suggest that language diversity correlates with reduced transmission rates.

Moro (2016) hypothesized that language diversity might have helped mitigate the problems of population growth when food supply and healthcare were limited. Combined with the more general idea that language diversity is shaped by physical and social factors external to language (Nettle, 1999; Lupyan and Dale, 2016), here we investigate potential correlations between language diversity and disease spread. We follow two complementary approaches. First, we compared the spread of COVID-19 in regions that show opposite linguistic landscapes. Following Kirkeby et al. (2017), we calculated the average daily COVID-19 transmission coefficient β for 101 countries using data from Johns Hopkins University. We then compared disease transmission in countries differing in linguistic diversity index (LDI) (from Eberhard et al. (2024)), while controlling for factors that could impact disease transmission: temperature, precipitation, elevation, latitude, human development index, road and population density, drawn from Global Data Lab, Wikipedia, and the World Bank. A linear regression was done between the difference in LDI (ΔLDI) and the difference in COVID-19 transmission rate $\Delta\beta$, with other factors as predictors. Second, to clarify the mechanism of language diversity on disease spread, we simulate the behavior of a pathogenic vector in the 6 society types discussed in (Trudgill, 2011), hypothesized to impact differentially on language features and language diversity. These 6 types vary in community size, network topology, and degree of contact. Specifically, we implemented an SEIR disease transmission model (Bjørnstad et al., 2020) for each of the 6 societal types. In our model, each type consisted of 4 societies, each with 4 communities. Size was modeled by changing the population number. Network topology and the

degree of contact were modeled by modifying the average number of people traveling from one community to another, or from one society to another, respectively (See Table 1). We implemented the model outlined in Lee and Jung (2015) with MATLAB (R2024b). To quantify disease spread, we assumed a single exposure case in community 1 from society 1 and measured the number of days for the number of susceptible people in each community to drop below 50%.

Table 1.: The six basic types of society simulated in the study

Type	Size	Network	Contact	Population	Within-society contact	Between-society contact
1	small	tight	low	1000	$[1, 1.5]$	$[0.1, 0.15]$
2	small	tight	high	1000	$[1, 1.5]$	$[1, 1.5]$
3	small	loose	low	1000	$[0.1, 0.15]$	$[0.1, 0.15]$
4	small	loose	high	1000	$[0.1, 0.15]$	$[1, 1.5]$
5	large	loose	low	4000	$[0.1, 0.15]$	$[0.1, 0.15]$
6	large	loose	high	4000	$[0.1, 0.15]$	$[1, 1.5]$

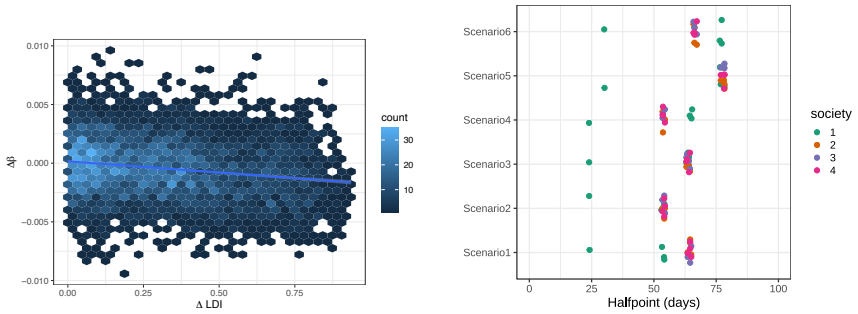


Figure 1.: **a**: Difference in transmission coefficient ($\Delta\beta$) vs. difference in Language Diversity Index (ΔLDI). **b**: Time (unit: days, x-axis) for the susceptible population to drop below 50% in each community across scenarios (y-axis). Communities from the same society are represented by the same color.

Our first analysis (Figure 1a) found a significant, negative effect of ΔLDI on $\Delta\beta$ (r =-2.665*10^{-4}, p = 0.021), suggesting language diversity reduces disease transmission independently of physical environment or social development. Our simulation (Figure 1b) showed in general, larger population sizes, looser social networks, and reduced contacts slow disease spread. Interestingly, scenarios 1 and 4/6 correspond to esoteric and exoteric languages, respectively (Wray and Grace, 2007; Chen et al., 2024). Esoteric languages are typically spoken in areas with high linguistic diversity. As shown, disease spread proceeds differently in both cases, with slower ingroup transmission in the esoteric scenario but slower intergroup transmission in the exoteric scenario. Overall, our findings give support to the view that a barrier effect to disease spread could be one of the factors promoting language diversity.

References

Bjørnstad, O. N., Shea, K., Krzywinski, M., and Altman, N. (2020). The seirs model for infectious disease dynamics. *Nature Methods*, 17(6):557–558.

Chen, S., Gil, D., Gaponov, S., Reifegerste, J., Yuditha, T., Tatarinova, T., Progovac, L., and Benítez-Burraco, A. (2024). Linguistic correlates of societal variation: A quantitative analysis. *PloS one*, 19(4):e0300838.

Eberhard, D. M., Simons, G. F., and Fennig, C. D., editors (2024). *Ethnologue: Languages of the World*. SIL International, Dallas, Texas, twenty-seventh edition edition. Online version: `http://www.ethnologue.com/`.

Kirkeby, C., Halasa, T., Gussmann, M., Toft, N., and Græsbøll, K. (2017). Methods for estimating disease transmission rates: Evaluating the precision of poisson regression and two novel methods. *Scientific reports*, 7(1):9496.

Lee, J. and Jung, E. (2015). A spatial–temporal transmission model and early intervention policies of 2009 a/h1n1 influenza in south korea. *Journal of theoretical biology*, 380:60–73.

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.

Moro, A. (2016). *Impossible languages*. MIT Press.

Nettle, D. (1999). *Linguistic diversity*. Oxford University Press.

Trudgill, P. (2011). *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford University Press, Oxford.

Wray, A. and Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3):543–578.

Linguistic experience and universal design of the auditory system in speech rhythm perception: implications for speech evolution

Mikhail Ordin [1,2], Leona Polyanskaya[3], David Maximiliano Gómez[4], and Arthur G. Samuel[5,6]

[1]Medical Faculty, Universidade de Coimbra, Portugal
[2]Coimbra Institute for Biomedical Imaging, Portugal
[3]Faculty of Psychology, Universidade de Coimbra, Portugal
[4]Universidad de O'Higgins, Chile
[5]Stony Brook University, New York, USA
[6]BCBL, Donostia, Spain

We investigated whether rhythm discrimination is mainly driven by the native language of the listener or by the fundamental design of the human auditory system and universal cognitive mechanisms shared by all people irrespective of rhythmic patterns in their native language. The results showed that rhythm change detection is a fundamental function of a processing system that relies on general auditory mechanisms and is not modulated by linguistic experience.

The existing literature is consistent with two plausible and reasonable hypotheses: Either (a) linguistic experience (primarily, one's native language) shapes rhythm processing, or (b) prosody in general (and rhythmic structures in particular) in natural languages is shaped by the general design of the auditory system, cognitive mechanisms, and neural physiology. On the one hand, rhythm perception is essential for speech processing and for language acquisition in infancy (Langus et al., 2018). Rhythmic patterns differ between languages (Gervain et al., 2008; Ramus & Mehler, 1999; White & Mattys, 2007) and thus individuals may differ in their experience with different rhythms. Non-native (Polyanskaya et al., 2017; Tajima et al., 1997) or pathological (Kent et al., 1989) rhythm affects speech accentedness and comprehensibility. These observations suggest that rhythmic patterns in speech might be processed via the phonological filter of the native language. In the other hand, rhythm perception relies on a fundamental design of mammalian auditory system (Gitza, 2011; Greenberg & Ainsworth, 2004; Hickok et al., 2015; Howard & Poeppel, 2012) that underlies rhythm discrimination by animals (Tincoff et al., 2005; Toro et al., 2003) and pre-linguistic babies (Nazzi & Ramus, 2003; Ramus et al., 1999). This mechanism is not unique to humans and is shared by all people irrespective of their native

language. We performed multiple experiments to pit these two hypotheses, both logically coherent and plausible according to prior empirical evidence, against one another.

We asked participants to listen to two continuous acoustic sequences and to determine whether their rhythms were the same or different (AX discrimination). Participants were native speakers of four languages with different rhythmic properties (Spanish and French – regular rhythm; English and German – irregular rhythm), to understand whether the predominant rhythmic patterns of a native language affect sensitivity, bias and reaction time (RT) in detecting rhythmic changes in linguistic (Experiment 2) and in non-linguistic (Experiments 1 and 2) acoustic sequences. Linguistic sequences were resynthesis of Welsh sentences using Italian phonemes and substituting those speech sounds that do not exist in Italian by the closet matching sounds of Italian language; Welsh intonation was lifted from the original sentences and imposed on resynthesized sentences. Non-linguistic stimuli were monotonous syllabic sequences randomly concatenated. Rhythm was manipulated by modulating variability in the vowel durations, higher variability means less regularity. Using a post-test naturalness evaluation experiment, participants rated linguistic stimuli as natural speech in a language they did not understand, similar to natural language resynthesized in the same manner. Non-linguistic stimuli were rated as less natural and not speech-like.

We examined sensitivity and bias measures, as well as RTs. We also computed Bayes factors in order to assess the effect of native language.

All listeners performed better (i.e., responded faster and manifested higher sensitivity and accuracy) when detecting the presence or absence of a rhythm change when the first stimulus in an AX test pair exhibited regular rhythm than when the first stimulus exhibited irregular rhythm. This result pattern was observed both on linguistic and non-linguistic stimuli and was not modulated by the native language of the participant.

We conclude that rhythmic cognition is based on general auditory and cognitive mechanisms and are not modulated by linguistic experience and are shared by all mammals. We suggest that the mechanisms are related to vocal learning, beat induction, and rhythmic entrainment (the ability to coordinate motor output with sensory input). The ability to discriminate rhythmic patterns in not only pre-requisite of speech development in ontogenesis, but also underlied speech emergence in phylogenesis of the homo genera. Irregular rhythmic patterns are marked, in a sense that any communicative system that exhibits irregular rhythm also exhibits regular rhythmic patterns, while there are communicative systems that only exhibit regular rhythms. Also, regular speech rhythms are more typologically common across languages, and the same markedness relations can be applied to speech rhythms. Regular rhythms allow better coupling between the

acoustic and neural oscillations and facilitate processing of the auditory input. Thus, switching from regular to irregular rhythm can be explained by expanding the repertoir of meanings to be expressed by prosodic means and facilitate the transition to the referentiality of the signals.

These results raise a series of further questions about rhythm perception and language evolution. We indeed suggested that the results are likely to be generalized across mammals (based on common neural-to-acoustic entrainment mechanism), yet we cannot say how far back in evolution we can go, whether it is generalizable across vertebrates (e.g., songbirds, who also exhibit remarkable vocal learning ability). These questions require dedicated empirical studies to be answered.

References

Gervain, J. Nespor, M., Mazuka, R., Horie, R., & Mehler J. (2008) Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. Cognitive Psychology 57(1), 56-74.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front. Psychol. 2:130.doi: 10.3389/fpsyg.2011.00130

Greenberg, S. & Ainsworth, W. (2004). Speech processing in the auditory system: An Overview. In S. Greenberg, W. Ainsworth, A. Popper, & R.Fay (Eds.). Speech Processing in the Auditory System (pp. 1-62). New York: Springer Verlag.

Hickok, G., Farahbod, H., & Saberi, K. (2015). The Rhythm of Perception: Acoustic Rhythmic Entrainment Induces Subsequent Perceptual Oscillation. Psychological Science 26(7), 1006–1013.

Howard, M., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustic but not comprehension. Journal of Neurophysiology 104, 2500-2511.

Kent, R.D., Weismer, G., Kent, J.F., & Rosenbek, J.C. (1989). Toward phonetic intelligibility testing in dysarthria. Journal of Speech and Hearing Disorders 54, 482–499.

Langus, A., Mehler, J., & Nespor, M. (2018). Rhythm in language acquisition. Neuroscience & Biobehavioral Reviews 81, 158-166.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. Speech Communication 41, 233-243.

Polyanskaya, L., Ordin, M., & Busa, M. (2017). Relative Salience of Speech Rhythm and Speech Rate on Perceived Foreign Accent in a Second Language. Language and Speech 60(3), 333-355.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. Journal of the Acoustical Society of America 105 (1), 512-521.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. Cognition 73(3), 265-292.
Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. Journal of Phonetics 25, 1-24.
Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: further tests with a non-human primate. Developmental Science (1), 26-35.
Toro, J.M., Trobalon, J.B, Sebastian-Galle, N (2003) The use of prosodic cues in language discrimination tasks by rats. Animal Cognition 6, 131–136.
White, L., & Mattys, S. (2007). Calibrating rhythm: First language and second language studies. Journal of Phonetics 35(4), 501-522.

Mindreading and language origins: An electroencephalogram study on the comprehension of basic ostensive communication

Francesco Ferretti[1,], Angelo D. Delliponti[2], Valentina Deriu[1], Alessandra Chiera[1], Daniela Altavilla[1], Serena Nicchiarelli[1], Sławomir Wacewicz[2] and Ines Adornetti*[1]

*Corresponding Author: ines.adornetti@uniroma3.it

[1]Cosmic Lab, Roma Tre University, Rome, Italy

[2]CLES, Nicolaus Copernicus University in Toruń, Toruń, Poland

We present an event-related potentials (ERPs) study on the time course of processing communicative and informative intentions expressed through a combination of eye contact and gestures. Our main finding from the ERPs analysis relates to the N170 component: it shows that the detection of both intentions occurs within the 200-millisecond window. This suggests that mindreading occurs through low-level rather than high-level cognitive processes. We address the evolutionary implications of this finding within the context of a deflationary perspective on ostensive communication.

A widely held view within language evolution research is that mindreading is a crucial cognitive prerequisite for the origin of language (e.g., Scott-Phillips, 2015; Tomasello, 2008). Such a view (mostly) rests on the ostensive-inferential model (Sperber & Wilson, 1995), which posits that human communication is a process of expression and recognition of intentions. Two types of intentions are in place in ostensive communication (OC): informative and communicative. *Informative intention* (II) is informing "by virtue of affecting the audience's mental state" (Scott-Phillips 2015, p. 58). *Communicative intention* (CI) is "the signaler's intention that the receiver recognizes that the signaler has an informative intention" (*ibidem*). Since mindreading is the cognitive system processing the mental states of others, and since intentions are mental states, such a system is a logical and temporal prerequisite of OC (Origgi & Sperber, 2000). A pivotal aspect of the ongoing debate focuses on the nature of the mindreading process underlying OC. In its classical formulation (Sperber & Wilson, 1995, 2002), the dual level of intention that characterizes OC is considered processed by high-level inferential mindreading for both CI and II processing (Sperber & Wilson, 2002), considered being an exclusive prerogative of adult humans, not available to human infants or nonhuman primates. In recent years, a deflationist perspective

on OC has emerged, challenging the prevailing classical view that OC is an all-or-nothing phenomenon (e.g., Moore, 2016, in press; Sperber & Wilson, 2024). This new approach associates basic forms of mindreading with basic forms of OC, thus opening the way to the view that basic forms of OC can be observed in both human infants and non-human primates. To adjudicate between these two perspectives, we present an ERPs study on the time course of processing of CI and II expressed through combination of eye contact and gestures.

Methods: 43 participants were recruited for the study. Their neural activity was recorded while they watched visual stories representing OC between two characters (Fig.1). Three conditions (randomly presented) are compared. *Congruent*: an interaction in which both the CI and II are satisfied (Fig. 1a). *Semi-congruent*: an interaction in which the II of the communicator but not her/his CI is satisfied (Fig. 1b). *Incongruent*: an interaction in which neither the CI nor the II are satisfied (Fig. 1c). There were 90 visual stories (30 for each condition) divided into 3 lists. The first 3 scenes were the same in each list (Fig. 1); the final scene changes for each list according to the condition. Each participant was administered a single list randomly assigned.

Results and discussion: A major finding emerging from the analysis of the ERPs relates to the amplitude of an early component, N170 (time window 100-210 ms): the incongruent condition had a more pronounced negative amplitude compared to the congruent condition ($p = .017$); no significant difference on this component emerged between the semi-congruent and congruent conditions, nor between the semi-congruent and incongruent conditions. Previous literature suggests that the N170 reflects neural activity associated with early-stage face processing (e.g., Eimer, 2011) and is crucially involved in mindreading processes (Ruzzante & Vaes, 2021). The absence of significant differences in the N170 response between congruent and semi-congruent conditions suggests that the processing observed in the semi-congruent condition is analogous to that of the congruent condition. This finding might be indicative that the elaboration of II occurs at an early stage, approximately 170 milliseconds after the initial presentation, in a manner analogous to CI, favoring conceptions of mindreading grounded in low-level rather than high-level cognitive processes. We will argue that such a finding supports a deflationary perspective on OC, i.e., a gradualist approach based on the possibility that even nonhuman animals can be ascribed forms of OC (albeit to varying degrees) (Moore, in press; Sperber & Wilson, 2024), which may have constituted the evolutionary roots of human communication.

Figure 1. The context of the story without the final events (represented in Fig. 1a 1b, 1c).

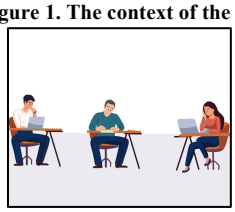


Fig. 1a. Congruent

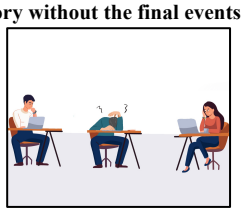


Fig. 1b. Semi-congruent

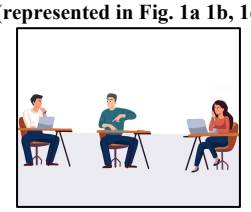
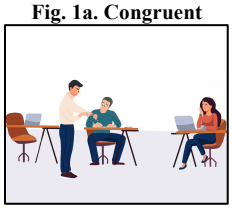
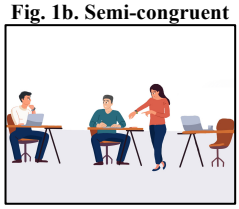
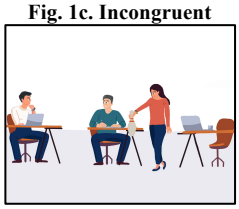


Fig. 1c. Incongruent







References

Eimer, M. (2011). The face-sensitivity of the n170 component. *Frontiers in human neuroscience*, *5*, 119.

Moore, R. (in press), Intentions in human and non-human great ape communication. In B. Geurts and R. Moore (Eds.), *Evolutionary pragmatics*. Oxford: Oxford University Press.

Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Animal Cognition,* 19(1), 223–231.

Origgi, G., & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the human mind: Language, modularity and social cognition* (pp. 140–169). Cambridge: Cambridge University Press.

Ruzzante, D., & Vaes, J. (2021). The timeline of mentalization: Distinguishing a two-phase process from mind detection to mind attribution. *Neuropsychologia*, *160*, 107983.

Scott-Phillips, T. C. (2015). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, *56*(1), 56-80.

Sperber, D., & Wilson, D. (2024). Rethinking ostensive communication in an evolutionary, comparative, and developmental perspective. *PsyArXiv: https://doi. org/10.31234/osf. io/zp3fx*.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press. (Original work published 1986).

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind reading. *Mind & Language, 17*(1–2), 3–23.

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

Mitteilungsbedürfnis: How the human drive to share mental states shaped the evolution of language, collaboration, and culture

Sabine Blütgen[*1], Jennifer Culbertson[1], and Simon Kirby[1]

[*]Corresponding Author: s.bluetgen@sms.ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, United Kingdom

Humans have a strong desire to share their thoughts and feelings with others (*Mitteilungsbedürfnis*, MtB) which is unique among our primate relatives, who do not usually give up information voluntarily. Why do humans have MtB? What pressures might have led to its evolution, and how is it advantageous to our species? Here, we review existing evidence and ideas from a range of disciplines and argue that MtB was crucial in the evolution of language, collaboration, and culture.

One of the main factors that sets human cognition and behaviour apart from that of most other animals is our intrinsic drive to communicate our thoughts, feelings, and experiences with others. Although there is evidence for non-human animals possessing rich mental lives (Rogers & Kaplan, 2024; Steiner, 2008), most seem to lack an intrinsic motivation to share them. This is true also for our closest primate relatives, who do not use communication in the cooperative, informative way that is typical for our species, and only share information when incentivised by external rewards (e.g. food) (Fitch, 2019; Tomasello & Call, 2019). Fitch (2010) calls this human drive *Mitteilungsbedürfnis* (MtB) and suggests that our overwhelming need to share may have created pressures for the evolution of a highly expressive symbolic communication system – i.e., language.

An equally unusual trait is our outstanding ability to collaborate with others – a skill which forms the essence of our culture, society, and humanity, and is one of the key aspects of our success as a species (Fitch, 2011; Smith, 2003). We propose that these two observations are linked, and that it is our MtB which allows us to collaborate on such an extraordinary level, based on three main factors:

I. MtB is a developmental pre-requisite for the acquisition of the ability to collaborate: sharing emotions and mental states with others is the first developmental step in children toward sharing goals and perceptions, which in turn enable the joint intentions and attention needed for collaboration (Tomasello et al., 2005).
II. Collaboration requires trust; honest information sharing increases trust among collaborators (Centorrino et al., 2015; Wu et al., 2016), and explicitly sharing our thoughts and feelings helps others to judge our intentions (Van Doorn et al., 2012). This is one way MtB helps us to

overcome collaboration challenges that are unscalable for most other animals, but comparatively easily solved by humans (Smith, 2003).

III. The free flow of information enabled by our MtB acts as a conduit in collaborative acts, and helps to coordinate actions via linguistic and non-linguistic communication (Dor, 2023; Smith, 2003).

Although Fitch proposed that MtB might have been a driving force in the evolution of language, it has only occasionally been mentioned in the literature, for example in work on the evolution of gestural communication (Fröhlich et al., 2019), syntax (Zuberbühler, 2020), and cognitive linguistics (Pleyer & Hartmann, 2014). Currently, the discussion of its existence, functions, and origins is scattered across disciplines, and there is no unified definition of what MtB actually entails. In addition, whilst there is some empirical evidence which shows that humans are behaviourally and neurally predisposed to share information with others (Tamir et al., 2015; Tamir & Mitchell, 2012), there is a lack of research testing inter-individual and cross-cultural variation in this proposed universal human trait.

Here, we present the first systematic review of MtB and its implications for the evolution of language and collaboration. Our main aims are as follows: (i) to review the many different ideas surrounding the human drive to communicate from the past 25 years of research in the cognitive sciences and beyond; (ii) to identify and unify common understandings and arguments in order to clearly define MtB; (iii) to explore its proposed connection to collaboration in human ontogeny, phylogeny, and ethology; (iv) to evaluate existing evidence for its existence; and, crucially, (v) to provide a basis to form hypotheses which allow us to test, among others, how MtB is expressed (including cross-culturally and inter-individually), how it develops in human ontogeny, how it interacts with collaborative success, and how it might have driven the evolution of language, collaboration, and culture.

Acknowledgements

SB is supported by funding from the School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

References

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2015). Honest signaling in trust interactions: Smiles rated as genuine induce trust and signal higher earning opportunities. *Evolution and Human Behavior*, *36*(1), 8–16. https://doi.org/10.1016/j.evolhumbehav.2014.08.001

Dor, D. (2023). Communication for collaborative computation: Two major transitions in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1872), 20210404. https://doi.org/10.1098/rstb.2021.0404

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511817779

Fitch, W. T. (2011). "Deep Homology" in the Biology and Evolution of Language. In A. M. D. Sciullo & C. Boeckx (Eds.), *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (pp. 135–166). Oxford University Press.

Fitch, W. T. (2019). The Comparative Approach to the Biology and Evolution of Language. In P. Hagoort (Ed.), *Human Language* (pp. 623–638). The MIT Press. https://doi.org/10.7551/mitpress/10841.003.0053

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809–1829. https://doi.org/10.1111/brv.12535

Pleyer, M., & Hartmann, S. (2014). A Matter of Perspective: Viewpoint Phenomena in the Evolution of Grammar. *The Evolution of Language*, 98–105. https://doi.org/10.1142/9789814603638_0011

Rogers, L. J., & Kaplan, G. (2024). Advances in understanding cognition in animals. In *Animals as Crime Victims* (pp. 53–66). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781802209884/book-part-9781802209884-12.xml

Smith, E. A. (2003). Human Cooperation: Perspectives from Behavioral Ecology. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 401–427).

Steiner, G. (2008). *Animals and the Moral Community: Mental Life, Moral Status, and Kinship*. Columbia University Press. https://doi.org/10.7312/stei14234

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences*, *109*(21), 8038–8043. https://doi.org/10.1073/pnas.1202129109

Tamir, D. I., Zaki, J., & Mitchell, J. P. (2015). Informing others is associated with behavioral and neural signatures of value. *Journal of Experimental Psychology: General*, *144*(6), 1114–1123. https://doi.org/10.1037/xge0000122

Tomasello, M., & Call, J. (2019). Thirty years of great ape gestures. *Animal Cognition*, *22*(4), 461–469. https://doi.org/10.1007/s10071-018-1167-1

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. https://doi.org/10.1017/S0140525X05000129

Van Doorn, E. A., Heerdink, M. W., & Van Kleef, G. A. (2012). Emotion and the construal of social situations: Inferences of cooperation versus competition from expressions of anger, happiness, and disappointment. *Cognition & Emotion*, *26*(3), 442–461. https://doi.org/10.1080/02699931.2011.648174

Wu, J., Balliet, D., & Van Lange, P. A. M. (2016). Gossip Versus Punishment: The Efficiency of Reputation to Promote and Maintain Cooperation. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep23919

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20190062. https://doi.org/10.1098/rstb.2019.0062

Navigation as the Evolutionary Origin for Recursive Processing in Human Cognition

Elizabeth Qing Zhang[1], Qi Wang[2], and Atonio Benítez-Burraco[*3]

[*]Corresponding Author: abenitez8@us.es

[1]Center of Language Evolution Studies, Nicolaus Copernicus University, Torun, Poland

[2]Department of Psychology, Sun Yat-Sen University, Guangzhou, China

[3]Department of Spanish, Linguistics, and Theory of Literature, University of Seville, Spain

In this work, we hypothesize that the navigation system supported by the hippocampus could have served as the evolutionary origin of domain general recursive computation in humans. Reviewing navigation and recursion abilities in invertebrates and vertebrates suggests that they correlate with each other. Neurobiological studies also show that the hippocampus plays a crucial role in both abilities. Developmental and clinical research supports our hypothesis as well. We further propose a key role of Human Self-Domestication in the hippocampal refinement.

Clarifying which aspects of our cognition can be regarded as specific to humans and how they could have evolved from precursors otherwise shared with other animals is a big concern for many disciplines from psychology to ethology to neuroscience. Recently, two lines of research have offered some promising insights into this complex issue. On the one hand, several studies have examined how concepts are computed by the brain, with a focus on the hippocampus and its role on navigation abilities (e.g. Courellis et al., 2024; Kazanina & Poeppel, 2023). On the other hand, ongoing research has identified a domain-general ability for processing recursive hierarchies, which manifests itself in language, music, mathematics, and other human-typical abilities (Dehaene et al., 2022; Fitch, 2014). This work aims to bridge these two disparate narratives of our cognitive uniqueness. We propose that the hippocampal mechanism for navigating the world could have served as the scaffolding for the evolution of our domain-general ability for computing recursive structures. Instead of supporting the view that humans independently evolved both a Broca's area-related mechanism for computing recursion and a hippocampal mechanism for computing a Language of Thought, we suggest that we improved the hippocampal navigation system (intra- and inter-hippocampal connections) and later applied to different domains, in line with the domain-general nature of most if not all representational and

computational brain devices, in line with Poeppel and Embick (2005) and many others.

For achieving this goal, we first review navigation abilities and recursion-like behaviors in other species, both invertebrates (ants, bees) and vertebrates (birds) including mammals (rodents) and primates (including humans). We focus on the navigation toolbox (Wiener et al., 2011; Jeffery et al., 2024) and Chomsky hierarchy (Heinz, 2016), respectively. We found that the more sophisticated navigation abilities one species has, the more complex the recursive-like behaviors it also exhibits. This supports the view that our advanced navigation abilities could have improved our recursive abilities, as found, most notably, in language).

To provide additional support to this view, we review selected neurobiological studies on navigation and recursive-like behaviors in animals including humans. We found enough evidence of a functional connection between, specifically, the hippocampus and brain areas known to be involved in (domain-general) recursion(-like) abilities.

Next, we provide further support to our view by reviewing evidence of i) the parallel improvement of navigation abilities and recursion abilities during child growth, and ii) the comorbidity of navigation impairment and recursion impairment in human prevalent cognitive disorders, such as Alzheimer's Disease (Coughlan et al., 2019; Ekstrom & Hill, 2023), particularly, in cases of hippocampal lesions.

Finally, we hypothesize that Human Self-Domestication (HSD), that is, our evolutionary trend towards more prosocial behaviors (this also impacting on our body and cognition), might have played a crucial role in this process by i) refining our hippocampal activity and ii) enhancing the connections between the hippocampus and other brain areas. Behavioral, neurobiological and genetic evidence support this view (see Benítez-Burraco, 2021 for details). To some extent, this process can be construed as extending our sophisticated navigation abilities from foraging to navigating a more complex social world.

In conclusion, our cognitive (non-)uniqueness has certainly bewildered cognitive scientists for decades. Our hypothesis provides a potential answer to the origins of one of our distinctive features, namely recursive processing, which might have evolved from navigation abilities with precursors in other species.

References

Benítez-Burraco, A. (2021). Mental time travel, language evolution, and human self-domestication. *Cognitive Processing*, *22*(2), 363–367. https://doi.org/10.1007/s10339-020-01005-2

Coughlan, G., Laczó, J., Hort, J., Minihane, A. M., & Hornberger, M. (2018). Spatial navigation deficits - overlooked cognitive marker for preclinical Alzheimer disease?. *Nature reviews. Neurology*, *14*(8), 496–506. https://doi.org/10.1038/s41582-018-0031-x

Courellis, H. S., Minxha, J., Cardenas, A. R., Kimmel, D. L., Reed, C. M., Valiante, T. A., Salzman, C. D., Mamelak, A. N., Fusi, S., & Rutishauser, U. (2024). Abstract representations emerge in human hippocampal neurons during inference. *Nature*, *632*(8026), 841–849. https://doi.org/10.1038/s41586-024-07799-x

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences*, *26*(9), 751-766.

Ekstrom, A. D., & Hill, P. F. (2023). Spatial navigation and memory: A review of the similarities and differences relevant to brain models and age. *Neuron, 111*(7), 1037–1049. https://doi.org/10.1016/j.neuron.2023.03.001

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, *11*(3), 329–364. https://doi.org/10.1016/j.plrev.2014.04.005

Heinz, J. (2016). Computational theories of learning and developmental psycholinguistics. In Lidz, Snyder & Pater (Ed.) *The Oxford Handbook of Developmental Linguistics*, Oxford University Press

Jeffery, K. J., Cheng, K., Newcombe, N. S., Bingman, V. P., & Menzel, R. (2024). Unpacking the navigation toolbox: insights from comparative cognition. *Proceedings of the Royal Society B: Biological Sciences*, *291*(2016). https://doi.org/10.1098/rspb.2023.1304

Kazanina, N., & Poeppel, D. (2023). The neural ingredients for a language of thought are available. In *Trends in Cognitive Sciences* (Vol. 27, Issue 11, pp. 996–1007). Elsevier Ltd. https://doi.org/10.1016/j.tics.2023.07.012

Laland, K., & Seed, A. (2021). Understanding Human Cognitive Uniqueness. *Annual Review of Psychology*, *72*(1), 689–716. https://doi.org/10.1146/annurev-psych-062220-051256

Poeppel, D & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In Anne Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*, 103–118. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Wiener, J., Shettleworth, S., Bingman, V. P., Cheng, K., Healy, S., Jacobs, L. F., ... & Newcombe, N. S. (2011). Animal navigation: a synthesis. In *Animal*

thinking: contemporary issues in comparative cognition (eds R Menzel, J Fischer), pp. 51–76. Cambridge, MA: MIT Press.

Neural evidence for a linguistic bias during speech segmentation in dogs

Kinga G. Tóth[*1,2,4]†, Boglárka Morvai[1]†, Marianna Boros[1], Dorottya S. Rácz[1,2], Ivaylo Iotchev[1], Kitti Szabó[1], Attila Andics[1,3],

[*]Corresponding Author: gtothk@gmail.com
[1]Neuroethology of Communication Lab, Department of Ethology, Eötvös Loránd University, Budapest, Hungary
[2]Doctoral School of Biology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary
[3]ELTE NAP Canine Research Group, Budapest, Hungary
[4]Cooperative Doctoral Program 2023, National Research, Development and Innovation Office, The Ministry of Culture and Innovation; Budapest, Hungary
† These authors contributed equally to this work.

During speech segmentation, humans rely more on the lexically loaded consonants than on the acoustically salient vowels. To investigate whether this consonant bias is human-unique, we compared consonant- vs. vowel-based speech segmentation in humans and family dogs, living in the speech-rich human environment. Neural evidence demonstrates that in dogs, like in humans, consonant-based segmentation is more efficient. Being exposed to speech may thus be sufficient for linguistic biases on statistical computations to emerge, even in an evolutionarily distant mammal.

Across languages, consonants carry more lexical information than vowels (McCarthy, 1979; Nespor et al., 2003). This may explain why in humans statistical computations underlying automatic word extraction from continuous speech streams (Saffran et al., 1996) rely more heavily on consonant than vowel patterns (Bonatti et al., 2005; Mehler et al., 2006; Toro et al., 2008a). But it is yet unknown whether consonant bias reflects a prelinguistic sound processing preference in humans. Such preference might have promoted the greater lexical informativity of future consonants. Alternatively, consonant bias could emerge experientially in language-exposed non-humans as well and may therefore be the consequence, rather than the cause, of the greater lexical load of consonants. To test this, we directly compared the EEG responses of humans (N=18) and dogs

(N=24) to continuous 7-min speech streams consisting of trisyllabic nonsense words defined by either consonant or vowel patterns (consonant-structured condition: e.g. muzita, mozite, mozüta; vowel-structured condition: e.g. kemuba, fezuba, kezuda). Neural entrainment to the word-level frequency (1.33 Hz), reflecting speech segmentation efficiency, was assessed by inter-trial coherence (ITC), a measure of phase consistency across trials at different frequencies (Batterink et al., 2017). We found stronger word-level (1.33Hz) neural entrainment for consonant- than vowel-structured speech streams not only in humans but also in dogs (Fig. 1.)

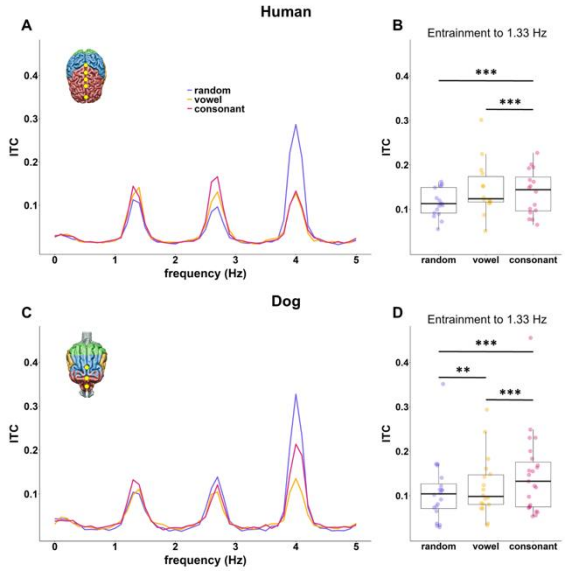


Fig 1. Electrode placements (A, C) **and results.** (A) Median ITC values for the human subjects in the frequency spectrum between 0 and 5 Hz. The presented frequency spectrum includes word frequency (1.33 Hz), its first harmonic (2.66 Hz), and syllable frequency (4 Hz). (B) ITC values at 1.33 Hz, for each condition in humans. (C) Median ITC values for the dog subjects. (D) ITC values at 1.33 Hz, for each condition in dogs.

These results provided the first neural evidence that consonant bias during speech segmentation is not unique to humans but is also present in dogs. Our findings suggest that exposure to speech may thus be sufficient for linguistic biases to emerge, even in evolutionarily distant, non-speaking species. Furthermore, this raises the possibility that in the early human brain certain speech processing mechanisms may have developed as a consequence of rapid evolutionary adaptation to the emerging, socially relevant stimulus, i.e., speech.

Acknowledgments

This project was funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant number 950159). KGT was supported by the Cooperative Doctoral Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund (KDP-2023). AA was supported by the National Brain Programme 3.0 of the Hungarian Academy of Sciences (NAP2022-I-3/2022).

References

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex 90*, 31–45.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler J. (2005). Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science 16(6),* 451–459.

McCarthy, J. J. (1979). *Formal Problems in Semitic Phonology and Morphology*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts (1979).

Mehler, J., Peña, M., Nespor, M., & Bonatti, L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior,* **42**(6), 846–854.

Nespor, M., Peña, M., & Mehler J. (2002). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* DOI:10.1418/10879

Saffran, J. R., Aslin, R. N., & Newport E. L. (1996). Statistical Learning by 8-Month-Old Infants. Science **13;274**(5294), 1926-8.

Toro, J. M., Nespor, M., Mehler J., & Bonatti, L. L. (2008). Finding Words and Rules in a Speech Stream: Functional Differences Between Vowels and Consonants. *Psychol Sci* **19**(2), 137-44.

Prominence in Protolanguage

T. Mark Ellison*

[1]SFB 1252, University of Cologne, Germany

We are prediction engines, and have been for longer than we have had language. Understanding language is a Bayesian inference task needing us to predict language production behaviour. To be efficient communicators, in terms of effort expended, producers must predict how understanders will interpret their productions. Linguistic prominence behaviour arises naturally as language producers use minimum effort to guide understanders. The prerequisites for prominence phenomena may have been already present in protolanguage.

1. Prediction and Prominence

Human beings are prediction engines, in fact, this may be a property of all living beings (Friston, 2010, 2012; Friston et al., 2023). Pickering and Garrod (2013) offer an account of how language understanders use their production capability to enhance their predictive capability during perception. In this paper, I describe model of communicative understanding and production relying on audience design and efficiency, enabled by prediction and social cognition. To the extent that these prerequisites are substantially present in great apes, the consequential features of communication likely arose early in protolanguage.

1.1. *Bayesian Interpretation*

The simplest model of communicative interpretation, relates the the prior expectation $P_U(m)$ by the understander U of the communicator referring to something m, to the likelihood $P(f|m)$ that they express it using a form f, given the prior likelihood $P(f)$ that f is used at all. Bayes theorem brings these into an expression for the likelihood $P(m|f)$ that the language producer intended meaning m (1).

$$P(m|f) \quad = \quad \frac{P(f|m)}{P(f)} P_U(m) \tag{1}$$

The expectation $P_U(m)$ of a referent m being referred to next can be equated with its *discourse prominence* von Heusinger and Schumacher (2019).

The remaining term on the right-hand side of (1), $\frac{P(f|m)}{P(f)}$, balances surprisal $1/P(f)$ of the form f with the aptness $P(f|m)$ of it expressing m. This combined term is called the *form prominence* of f. Note that the form prominence cannot be determined apart from its aptness for expressing a meaning.

1.2. *Audience Design*

Producers communicate better if they model the recipient's interpretation process. Avoiding uncertainty in interpretation, they may maintain a threshold ϑ level of *interpretative confidence* $P(m|f)$ in their intended meaning m. To do this, they need to model the understander's expectations $P_U(m)$ about what meanings are coming up next, i.e. their listeners discourse prominence levels. Then they should choose a form with the right level of specificity to ensure understander confidence at or above the threshold (2) for the correct meaning.

$$\vartheta \leq \frac{P(f|m)}{P(f)} P_U(m) \tag{2}$$

At the same time, Friston's notion of uncertainty minimisation suggests that language producers will reduce their uncertainty in their own choices, i.e. choosing the most likely form which satisfies the threshold. This will be an expression which minimises the *form prominence*.

1.3. *Implications*

Form prominence minimisation by producers means that as far as the options available in the language allow, the inequality in (2) will approach equality. Thus form prominence will trade off against discourse prominence: unexpected meanings will be expressed with less common, more precise forms. This behaviour is found at many levels of linguistic structure, e.g. syntax (Jaeger & Levy, 2006), lexical choice (Mahowald, Fedorenko, Piantadosi, & Gibson, 2013), phonetics (Aylett & Turk, 2004). As an example, pronouns become key tools in communicative efficiency. Referents with high discourse prominence is efficient with a high-frequency, ambiguous form (i.e. one with low form prominence).

2. Prominence in Protolanguage

The above model of communication requires only that the communication producers and interpreters are both predictive agents, able to model each others' behaviour. Given that great apes have substantial social cognition, including the ability to understand that other's access to information is different to their own (Tomasello, 2023), it is reasonable to expect that as our ancestors moved into the proto-language stage, the implications of prominence also applied. Consequently, many prominence-phenomena, such as pronominalisation, may have arisen early in the proto-language stage.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-Project-ID 281511265-SFB 1252 Prominence in Language - in the project C09 “Prominence and Probabilistic Modelling” at the University of Cologne, Institute for Linguistics - Phonetics.

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.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, *11*(2), 127–138.

Friston, K. (2012). A free energy principle for biological systems. *Entropy*, *14*(11), 2100–2121.

Friston, K., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G. A., & Parr, T. (2023). The free energy principle made simpler but not too simple. *Physics Reports*, *1024*, 1–29.

Jaeger, T., & Levy, R. (2006). Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems*, *19*.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313–318.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347.

Tomasello, M. (2023). Social cognition and metacognition in great apes: a theory. *Animal Cognition*, *26*(1), 25–35.

von Heusinger, K., & Schumacher, P. B. (2019). Discourse prominence: Definition and application. *Journal of Pragmatics*, *154*, 117–127.

Reconstructing linguistic proto-forms using phylogenetically-aware Neural Networks

Frederik Hartmann[*1]

[1]Linguistics Department, University of North Texas, Denton, USA

Reconstructing proto-forms of lexical items using computational models is a notoriously difficult task. Although attempts have been made to reconstruct proto-forms from a set of descendants using Neural Networks, these models do not account for phylogenetic information of the language family in question. The paper at hand tests Recurrent and Transformer Neural Network architectures that can take the linguistic family trees into account. The results show that incorporating the family tree can give a minor boost.

The computational reconstruction of proto-forms of lexical items has been attempted in recent years with various different methods and model types. Chief among those, there has been a rise in Neural Network approaches (Hartmann, 2021), (Fourrier, 2022), (Lu, Xie, & Mortensen, 2024), (Fourrier & Sagot, 2020). This task is notoriously difficult since it combines a small-data problem with under-labeled data, with very little data for training and testing. Most studies focus on reconstructing an ancestral form from a set of descendant items without adding information about the family tree. However, information on the relatedness of the languages the lexical items were drawn from, could enable the Neural Network to detect language-specific sound changes and account for those. The paper at hand tests a Neural Network model that can take into account the tree structure of a reconstruction problem to gauge whether such an architecture is feasible.

Specifically, a model was created that can take as input phonetically transcribed descendant words split into individual segments, a language family tree in the Newick format, and branch lengths for the associated tree. The model was first trained on a simulated dataset obtained through the following process. The *Indo-European lexical cognate data* (IELex) (Dunn & Tresoldi, 2022) database was taken as a basis for a cognate-annotated lexical dataset. To be able to accurately check the accuracy of the model predictions, sound changes were simulated in a language family over time to have full control over the linguistic patterns in the dataset. For this, we can create a random family tree with varying branch lengths and six terminal nodes using the R package *ape* (Paradis & Schliep, 2019). Simultaneously, the Python library *lingpy* (List & Forkel, 2021) was used to align the words in the dataset by segments for each individual cognate. Merging those

by bigrams to capture the immediate phonological environment of each sound provided a list of existing sound variations and their immediate environment in the languages in the dataset. To simulate the sound changes in a tree, the original lexical dataset was taken and assigned as the 'ancestor lexicon' to the root of the tree. For each edge in the tree, a number of bigram pairs from the bigram variation list were randomly selected, proportional to the edge length, with a maximum of 20 changes per tree, and applied to the lexicon. This was done iteratively for every branch until an ancestor lexicon and six descendant lexicons, which have gone through a series of consecutive regular sound changes, remained. Lastly, those cognates where less than 80 percent of descendants have changed to avoid having a large number of unchanged words in the dataset were filtered out. This was repeated 100 times, which gives 100 simulated trees, each with approximately 2,500 cognate sets. This procedure has the advantage that the final dataset consists of a known ancestor lexicon and a known tree structure, so the model results can be directly compared to the exact response (however, see discussion on the drawbacks below).

The models themselves are custom-built graph-based LSTMs and Graph-Transformer NNs. Both take the descendant words, the tree structure as an adjacency matrix, and a branch lengths matrix as inputs. Both networks first embed the characters in each word before The Graph-LSTM model does this by having the individual LSTM cells move over each cognate individually, sharing information from the cell and hidden states with LSTM cells of words at adjacent nodes in the tree. The Graph-Transformer, on the other hand, derives a soft attention mask from the adjacency matrices and thus attends more to segments in words that are connected nodes in the tree. Both networks receive as input the descendant words and predict both words at the descendant and ancestor nodes. Having the network predict all words at all nodes, even those it has received as input, improves training, while accuracy is only calculated on the unseen ancestor nodes.

Both models reconstruct on average 94% of all segments in the ancestor words correctly, on a validation set with 10 unseen trees (10% of trees in the dataset, random baseline: $\lesssim$ 1%). When running the models on the data without the tree-attention / tree-memory architecture, accuracy drops to 76% for the transformer model and to 42% for the LSTM model. This shows that incorporating phylogenetic graph structures into the attention mechanism of a transformer or the memory of parallel LSTM cells enables the models to learn structured relationships between cognates and to apply family-independent reconstructions to unseen trees.

References

Dunn, M., & Tresoldi, T. (2022, March). *evotext/ielex-data-and-tree: IELex data and tree (2022/03/28).* Zenodo.

Fourrier, C. (2022). *Neural Approaches to Historical Words Reconstruction.*

Theses, Université Paris sciences et lettres.

Fourrier, C., & Sagot, B. (2020). Comparing statistical and neural models for learning sound correspondences. In R. Sprugnoli & M. Passarotti (Eds.), *Proceedings of lt4hala 2020 - 1st workshop on language technologies for historical and ancient languages* (pp. 79–83). Marseille, France: European Language Resources Association (ELRA).

Hartmann, F. (2021). The phonetic value of the proto-indo-european laryngeals: A computational study using deep neural networks. *Indo-European Linguistics*, *9*(1), 26–84.

List, J.-M., & Forkel, R. (2021, July). *Lingpy. a python library for historical linguistics.* Zenodo.

Lu, L., Xie, P., & Mortensen, D. R. (2024). Semisupervised neural proto-language reconstruction. *arXiv preprint arXiv:2406.05930*.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526-528.

Revising Hockett's 'Design Features of Language'

Michael Pleyer[*1], Marcus Perlman[2], Gary Lupyan[3], Koen de Reus[4,5,6] and Limor Raviv[6,7]

[*]Corresponding Author: pleyer@umk.pl

[1]Center for Language Evolution Studies, Nicolaus Copernicus University in Toruń, Toruń, Poland

[2]Department of Linguistics and Communication, University of Birmingham, Birmingham, UK

[3]Department of Psychology, University of Wisconsin-Madison, Madison, USA

[4]Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Nijmegen, NL

[5]Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, Belgium

[6]Donders Centre for Cognition, Radboud University, Nijmegen, NL

[7]LEADS Group, Max Planck Institute for Psycholinguistics, Nijmegen, NL

Hockett's design features of language have been highly influential in language evolution research and comparative work on animal communication. However, a modern understanding of language necessitates a critical rethinking of this framework. Here we focus on three important domains that have wide-ranging implications for Hockett's design features - (i) multimodality and semiotic diversity; (ii) the functions of language, and (iii) language as an adaptive system - and show how these can better reflect the current research landscape and facilitate cross-species comparisons.

Language is seen as one of, if not *the*, most defining traits of our species. But what is language, and what makes it so unique? In 1960, Charles Hockett published "The origin of speech", enumerating 13 design features that are allegedly shared across human languages, and when taken together, distinguish language from other communication systems (Hockett, 1960). This foundational work marked a renewed evolutionary approach to the study of language, and has been highly influential. Hockett's publications on the topic (e.g. Hockett 1960; 1963) have been cited thousands of times, and his design features remain a mainstay of class syllabi, and are referenced in many linguistics and cognitive science textbooks (cf. Wacewicz et al. 2022) as well as popular press

publications (e.g. Johansson 2021). Moreover, for generations of scientists, Hockett's design features have been the default go-to model for evaluating and comparing non-human animal communication systems with human language.
However, in the roughly 65 years since they were first proposed, these features have been criticized (e.g. Oller 2004), and even called a "non-starter" (Wacewicz & Żywiczyński 2015). Extending on these important critiques, here we argue that Hockett's design features need to be thoroughly revised and updated to bring them in line with advances in linguistics, cognitive science, and animal communication and cognition to facilitate productive inquiry into language evolution. Specifically, we focus on three broad domains that characterize our modern understanding of language, and that each highlight the need for a fundamental re-evaluation and reconceptualization of Hockett's design features: (i) **multimodality and semiotic diversity**; (ii) **the functions of language**, and (iii) **language as an adaptive system**. For example, while one of Hockett's design features is the *vocal-auditory channel*, modern research has shown that language is fundamentally multimodal (e.g. Cohn & Schilperoord, 2024; Özyürek, 2021). Language exhibits modal flexibility and can be realised in different modalities (e.g., sign language; Stokoe, 2005). Even speech is tightly integrated with semiotic channels such as co-speech gesture, gaze, and facial expressions (Levinson & Holler 2014). Similarly, many animal species communicate using multimodal signals (Zhang & Pleyer, 2024), highlighting the prevalence of multimodality across communication systems. Another example is Hockett's feature of *specialization,* arguing that the function of language is the transfer of semantic information using a discrete "code". However, modern research shows that language serves multiple functions, such as social signaling (Smaldino & Turner 2022) and cognitive augmentation (Lupyan 2012), and its meaning-making relies heavily on pragmatics, ostension, and inference (Heintz & Scott-Phillips 2023). In fact, ostension is central for characterising human language as it enables interlocutors to use any behaviour and turn it into a communicative act. We therefore stress the importance of ostensive-inferential communication for revising the design features of language. In addition, while Hockett conceived language as characterised by a set of static features, modern research views language as a fundamentally dynamic and adaptive system (Beckner et al. 2009), which is continuously changing. In this sense, the design feature of *cultural transmission* is a unique process that in itself drives the creation of other design features such as *arbitrariness, semanticity* and *duality of patterning* via interaction and transmission, as shown in experimental research on the cultural evolution of language (Motamedi et al. 2019; Kirby et al. 2015; Raviv et al. 2019). We offer a comprehensive re-evaluation of Hockett's design features in light of these three overarching principles and show that this new framework not only better reflects the current research landscape, but also serves as an improved roadmap for future comparative work in animal communication and cognition.

References

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language Is a Complex Adaptive System: Position Paper. Language Learning, 59, 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x

Cohn, N., & Schilperoord, J. (2024). *A Multimodal Language Faculty: A Cognitive Framework for Human Communication*. Bloomsbury Publishing.

Heintz, C., & Scott-Phillips, T. (2023). Expression unleashed: The evolutionary and cognitive foundations of human communication. Behavioral and Brain Sciences, 46, e1. https://doi.org/10.1017/S0140525X22000012

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88-97.

Hockett, C. F. (1963). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals of Language*, 2, (pp. 1–29). MIT Press.

Johansson, S. (2021). *The Dawn of Language: The story of how we came to talk* (F. Perry, Trans.). MacLehose Press.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102. https://doi.org/10.1016/j.cognition.2015.03.016

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130302. . https://doi.org/10.1098/rstb.2013.0302

Lupyan, G. (2012). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 57, pp. 255–297). Academic Press.

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.

Oller, D. K. (2004). Underpinnings for a theory of communicative evolution. In D. K. Oller & U. Griebel (Eds.), Evolution of communication systems: A comparative approach (pp. 49–65). MIT Press.

Özyürek, A. (2021). Considering the Nature of Multimodal Language from a Crosslinguistic Perspective. *Journal of Cognition*, 4(1), 42. https://doi.org/10.5334/joc.165

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. Cognition, 182, 151–164. https://doi.org/10.1016/j.cognition.2018.09.010

Smaldino, P. E., & Turner, M. A. (2022). Covert signaling is an adaptive communication strategy in diverse populations. *Psychological Review*, 129(4), 812.

Stokoe, W. C., Jr. (2005). Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *The Journal of Deaf*

Studies and Deaf Education, 10(1), 3–37. https://doi.org/10.1093/deafed/eni001

Wacewicz, S., & Żywiczyński, P. (2015). Language Evolution: Why Hockett's Design Features are a Non-Starter. *Biosemiotics*, 8(1), 29–46. https://doi.org/10.1007/s12304-014-9203-2

Wacewicz, S., Pleyer, M., Szczepańska, A., Ewa Poniewierska, A., & Żywiczyński, P. (2022). The representation of animal communication and language evolution in introductory linguistics textbooks. Journal of Language Evolution, 7(2), 147–165. https://doi.org/10.1093/jole/lzac010

Zhang, E. Q., & Pleyer, M. (2024). Toward Interdisciplinary Integration in the Study of Comparative Cognition: Insights from Studying the Evolution of Multimodal Communication. *Comparative Cognition & Behavior Reviews*, 19, 85–90. https://doi.org/10.3819/CCBR.2024.190017

Sadat Tawaher Sign Language: Emergence, development, & typological novelty

Seyyed Hatam Tamimi Sa'd[*1], Ronnie B. Wilbur[2]

[*]Corresponding Author: shtamimi90@gmail.com
[1]Indo-American Center (IAC), Chicago, USA
[2]Department of Linguistics, Purdue University, West Lafayette, USA

Approximately sixty years ago, a sign language began to emerge within a single household in a small Iranian village, following a man's sudden, unexpected, and complete loss of hearing. Developed by hearing family members, this sign language exhibits grammatical features that differ from the surrounding spoken Arabic yet resemble those found in established sign languages. This sign language constitutes a new category of language emergence—one that neither fits within the current sign language typology nor has been previously reported.

Abstract

In this study, we present a first-hand historical account dating back approximately six decades ago, detailing a remarkable event that led to the emergence of a new sign language (SL): *Sadat Tawaher Sign Language (STSL)*. Specifically, we pursue a two-fold objective regarding STSL: (i) to examine its syntactic divergences from the surrounding spoken language, and (ii) to determine its placement within the broader typology of sign languages. STSL (in Arabic, *Lughat Isharat Sadat al-Tawaher*, *LIST*) emerged naturally around sixty years within a single household in the small village of *Sadat Tawaher* (Fig. 1), located in southwestern Iran, following the sudden and complete loss of hearing of a 20-year-old man named *Hanash* (Fig. 2). At the time, Hanash had no reading literacy, and there was no access to deaf education, leaving 'gesture' as the sole means of communication between him and his family and friends. What distinguishes this case is the subsequent development: in the absence of deaf education, prior knowledge of, or exposure to any SL, the deaf individual's family—all hearing native speakers of Khuzestani Arabic (KhA)—ingeniously developed an elaborate gestural system to communicate with Hanash. Over the past six decades, STSL has gradually evolved into a fully developed language, effectively serving a wide range of social and personal communication needs. Furthermore, although it is well-established that the surrounding spoken language typically influences

the development of a nearby SL (Meir, Sandler, Padden, & Aronoff, 2010), STSL is not *Manually Coded KhA*, but rather a typologically distinct language. Data were collected from 14 native signers (4 females and 10 males), ranging in age from 7 to 56, using a variety of elicitation techniques—including isolated signed sentence production, free storytelling, picture and video description tasks, and grammaticality judgment tests. The findings reveal that, whereas negative markers and wh-words appear in preverbal and clause-initial positions respectively in KhA (1a, 2a), they predominantly occur in sentence-final position in STSL (1b, 2b). Regarding the signing community, STSL has been acquired across three generations, encompassing family members, including the deaf individual's siblings, children, and grandchildren, as well as non-family members such as close friends and neighbors (Figures 2, 3, & 4). The signers range in age from 7 to 65 years old. Most family members acquired STSL alongside Khuzestani Arabic (KhA) as their mother tongues, thereby qualifying as bimodal bilinguals, i.e., competent in both a spoken and a signed language. STSL signers not only produce signs concurrently with speech but also regularly employ two distinct word orders simultaneously with minimal effort (1–2). Furthermore, STSL demonstrates not only syntactic differences from KhA but also notable similarities to other fully developed sign languages, such as the use of sentence-final negation and wh-signs (Zeshan, 2006). Nevertheless, it does not fit within the current sign language typology (e.g., de Vos & Pfau, 2015; Padden, 2011; Pfau, 2012; Zeshan, 2008) as it was uniquely developed by hearing individuals to facilitate communication with a late-deafened adult. Moreover, the deaf person did not sign himself as he retained full speaking abilities and therefore did not play a direct role in the creation of STSL. Notably, Hanash was the only deaf person in the village of Sadat Tawaher. As such, STSL does not precisely fit the definition of a homesign, deaf community / shared / emerging / village / or alternate / secondary SL (e.g., Goldin-Meadow, 2012; Meir et al., 2010; Pfau, 2012); rather, it represents a unique communication system that appears to constitute its own, previously undocumented category—one that has not yet been addressed in the existing literature and warrants further scholarly investigation.

References

de Vos, C., & Pfau, R. (2015). Sign language typology. *Annual Review of Linguistics*, *1*, 265-288.

Goldin-Meadow, S. (2012). Homesign. In R. Pfau, M. Steinbach, & B. Woll (eds.), *Sign Language* (pp. 601-625). Berlin / Boston: De Gruyter Mouton.

Meir, I., Sandler, W., Padden, C., & Aronoff, M. (2010). Emerging sign languages. In M. Marschark & P. E. Spencer (eds.), *Oxford handbook of deaf studies, language, and education* (pp. 267-280), Vol. 2. Oxford: Oxford University Press.

Padden, C. A. (2011). Sign language geography. In G. Mathur, & D. J. Napoli (Eds.), *Deaf around the world* (pp. 19-37). Oxford: Oxford University Press.

Pfau, R. (2012). Manual communication systems. In R. Pfau, M. Steinbach, & B. Woll (eds.), *Sign Language* (pp. 513-551). Berlin / Boston: De Gruyter Mouton.

Zeshan, U. (2006). Negative and interrogative constructions in sign languages. In U. Zeshan (ed.), *Interrogative and negative constructions in sign languages* (pp. 28-68). The Netherlands: Ishara Press.

Zeshan, U. (2008). Roots, leaves and branches. In R. M. de Quadros (ed.), *Sign languages* (pp. 671-695). Petrópolis: Arara Azul.

Examples:

(1)	a. KhA:	əl-wadəm	**ma**-ħəṣdæt	ʃələb		
		the-people	neg-harvest.perf	rice		(S-Neg-VO)
	b. STSL:	PEOPLE	HARVEST	RICE	**NOT**	(SVO-Neg)
		‘The people didn’t harvest the rice.’				
(2)	a. KhA:	əl-wælæd	**wiyæn**	ħæṭṭ	əl-qælæm	
		the-boy	where	put.3sm.perf	the-pen	(S-Wh-VO)
	b. STSL:	BOY	PEN	PUT	**WHERE**	(SOV-Wh)
		‘Where did the boy put the pen?’				

Figures:

Fig. 1. Sadat Tawaher

Fig. 2. Second generation signers

Fig. 3. Hanash & some signers (right to left)

Fig. 4. Third generation signers

Searching for the origins of recursive thinking via the complexity of stone toolmaking

Antoine Muller*[1], Larissa Mendoza Straffon[1,2], Ceri Shipton[3], and Francesco d'Errico[1,4]

*Corresponding Author: antoine.muller@uib.no

[1] SFF Centre for Early Sapiens Behaviour (SapienCE), University of Bergen, Bergen, Norway

[2] Department of Psychosocial Science, University of Bergen, Norway

[3] Institute of Archaeology, University College London, London, UK

[4] Université de Bordeaux, CNRS, Ministère de la Culture, PACEA, UMR5199, 33615 Pessac, France

In this contribution we attempt to identify recursive patterns in stone toolmaking behaviours, reconstructed from archaeological and experimental evidence. We wish to clarify the extent to which elements from the archaeological record meet linguistic conceptualizations of recursion. With case studies from throughout the Stone Age, we find evidence that the sequences of gestures, tasks, and goals involved in some stone tool technologies are hierarchical and fractal, but not necessarily recursive sensu stricto. We explore some possible candidates of elements of stone toolmaking that might more closely fit linguistic recursion. Translating linguistic recursion to the technological domain remains a key challenge for archaeologists seeking traces of the evolutionary history of language through material remains.

1. Abstract

Stone toolmaking, observed archaeologically from at least 3.3 million years ago, involves fashioning implements out of stone with different sequences of gestures, tasks, and goals. Many of these sequences are hierarchically complex (Mahaney 2014; Moore 2011; Muller *et al.* 2017; Shipton *et al.* 2013; Stout 2011), and potentially recursive too. Here, we draw on experimental and archaeological evidence to explore complexities of stone toolmaking behaviours, with an eye towards using the archaeological record to test the gestural origins of language hypothesis (Arbib 2012; Arbib *et al.* 2008; Corballis 2003; Hewes 1973; Meguerditchian 2022; Pulvermüller 2014; Rizzolatti & Arbib 1998). There is abundant proxy archaeological evidence, especially via symbolic material culture, for the possibility of a deep antiquity of language (Barham & Everett 2021; Bender *et al.* 2024; d'Errico *et al.* 2003; d'Errico & Vanhaeren 2009;

Henshilwood *et al.* 2001; Henshilwood & Dubreuil 2011), potentially supported by theories of grammaticalisation (Heine & Kuteva 2007). Recursive reasoning, sensu lato, has even been linked to toolmaking in the archaeological record (Haidle 2010; Hoffecker 2007; Lombard & Haidle 2012; Pelegrin 2009; Shipton *et al.* 2013; Shipton 2024; Stout 2011; Stout & Chaminade 2009; Wadley 2010). Yet, some of these behaviours might better be described as repetitive, nested, and fractal, but not necessarily recursive sensu stricto. Even for very complex stone tool technologies, it can be difficult to contradict the null hypothesis that they could have been parsed iteratively with a complex, but not recursive, chain of steps.

In an attempt to trace the chronology of recursive reasoning we draw on both experimental and archaeological evidence. We reconstruct the minutia of the actions recorded from videotaped stone toolmaking experiments, in which expert stone toolmakers manufactured Early Stone Age bifaces and Middle Stone Age cores with controlled raw materials and hammerstones (following methods outlined in Muller *et al.* 2017). The resultant flakes were collected in sequence and analysed, and the footage was used to time the duration of gestures and tasks. These data are used to reconstruct the size of flakes produced throughout the sequence, as well as the time spent focussed on different goals and sub-goals, helping to quantify the complexity and structure of the toolmaking sequences. We seek to ground truth these experimental observations with analyses of 3D scans of bifaces and cores from African Stone Age sites. These scans are used to precisely measure the size of flake scars preserved on the surface of these archaeological stone artefacts. Doing so quantifies the patterning in the size and location of scars, providing insight into the stages, goals, and sub-goals involved in these toolmaking sequences. We explore instances of biface and core manufacture where skewed scar proportions reveal similar sets of actions enacted on smaller and smaller scales.

We first quantify the ways in which some of these sequences are repetitive and fractal, but not necessarily recursive sensu stricto. Next, we ask whether the gestures and goals of some of these stone toolmaking sequences are recursive in a linguistic sense? In particular, we highlight specific sub-sequences of some technologies where stone toolmakers make counter-intuitive detours to solve unexpected problems as they arrive, before returning to the overarching goal. Others have shown that imparting the significance of these actions to observers requires careful gestures (Lycett & Eren, 2019; Shipton, 2024), possibly even pantomime (see Abramova 2018; Arbib 2002). Using the filmed experimental sequences, we explore how these diversions create similar nested sequences on a

smaller scale, highlighting processes like platform preparation, where intricate steps of the toolmaking process are scaled-down several times to achieve successive sub-goals.

With these case studies from Early Stone Age handaxes and Middle Stone Age Levallois cores, we chart the evolution in the complexity of these stone toolmaking sequences throughout the *Homo* lineage. We explore the challenges in translating recursion to the behavioural and technological domain and the extent to which these sequences of toolmaking gestures can be thought of as recursive, if at all.

References

Abramova, E. (2018). The role of pantomime in gestural language evolution, its cognitive bases and an alternative. *Journal of Language Evolution*, *3*(1), 26–40. https://doi.org/10.1093/jole/lzx021

Arbib, M. A. (2002). The mirror system, imitation, and the evolution of language. In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in Animals and Artifacts* (pp. 229–280). The MIT Press.

Arbib, M. A. (2012). *How the Brain Got Language: The Mirror System Hypothesis*. Oxford University Press.

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Current Anthropology*, *49*(6), 1053–1076. https://doi.org/10.1086/593015

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. https://doi.org/10.1007/s10816-020-09480-9

Bender, A., Straffon, L. M., Gatewood, J. B., & Beller, S. (2024). The dual role of culture for reconstructing early sapiens cognition. *Psychological Review*, *131*(6), 1411–1434. https://doi.org/10.1037/rev0000462

Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, *26*(2), 199–208. https://doi.org/10.1017/S0140525X03000062

d'Errico, F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A. M., Soressi, M., Bresson, F., Maureille, B., Nowell, A., Lakarra, J., Backwell, L., & Julien, M. (2003). Archaeological evidence for the emergence of language, symbolism, and music—An alternative multidisciplinary perspective. *Journal of World Prehistory*, *17*(1), 1–70. https://doi.org/10.1023/A:1023980201043

d'Errico, F., & Vanhaeren, M. (2009). Earliest personal ornaments and their significance for the origin of language debate. In R. Botha & C. Knight (Eds.), *The Cradle of Human Language* (pp. 16–40). Oxford University Press.

Haidle, M. N. (2010). Working-memory capacity and the evolution of modern cognitive potential: Implications from animal and early human tool use. *Current Anthropology*, *51*(S1), S149–S166. https://doi.org/10.1086/650295

Heine, B., & Kuteva, T. (2007). *The Genesis of Grammar: A Reconstruction*. Oxford University Press.

Henshilwood, C. S., D'errico, F., Marean, C. W., Milo, R. G., & Yates, R. (2001). An early bone tool industry from the Middle Stone Age at Blombos Cave, South Africa: Implications for the origins of modern human behaviour, symbolism and language. *Journal of Human Evolution*, *41*(6), 631–678. https://doi.org/10.1006/jhev.2001.0515

Henshilwood, C. S., & Dubreuil, B. (2011). The Still Bay and Howiesons Poort, 77-59 ka: Symbolic material culture and the evolution of the mind during the African Middle Stone Age. *Current Anthropology*, *52*(3), 361–400. https://doi.org/10.1086/660022

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology*, *14*(1/2), 5–24. https://doi.org/10.1086/201401

Hoffecker, J. F. (2007). Representation and recursion in the archaeological record. *Journal of Archaeological Method and Theory*, *14*(4), 359–387. https://doi.org/10.1007/s10816-007-9041-5

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. https://doi.org/10.1017/S095977431200025X

Lycett, S. J., & Eren, M. I. (2019). Built-in misdirection: On the difficulties of learning to knap. *Lithic Technology*, *44*(1), 8–21. https://doi.org/10.1080/01977261.2018.1539322

Mahaney, R. A. (2014). Exploring the complexity and structure of Acheulean stoneknapping in relation to natural language. *PaleoAnthropology*, *2014*, 586–606.

Meguerditchian, A. (2022). On the gestural origins of language: What baboons' gestures and brain have told us after 15 years of research. *Ethology Ecology & Evolution*, *34*(3), 288–302. https://doi.org/10.1080/03949370.2022.2044388

Moore, M. W. (2011). The design space of stone flaking: Implications for cognitive evolution. *World Archaeology*, *43*(4), 702–715. https://doi.org/10.1080/00438243.2011.624778

Muller, A., Clarkson, C., & Shipton, C. (2017). Measuring behavioural and cognitive complexity in lithic technology throughout human evolution. *Journal of Anthropological Archaeology*, *48*, 166–180. https://doi.org/10.1016/j.jaa.2017.07.006

Pelegrin, J. (2009). Cognition and the emergence of language: A contribution from lithic technology. In S. A. de Beaune, F. L. Coolidge, & T. Wynn (Eds.),

Cognitive Archaeology and Human Evolution (pp. 95–108). Cambridge University Press.

Pulvermüller, F. (2014). The syntax of action. *Trends in Cognitive Sciences*, *18*(5), 219–220. https://doi.org/10.1016/j.tics.2014.01.001

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194. https://doi.org/10.1016/S0166-2236(98)01260-0

Shipton, C. (2024). Recursive narrative and the Acheulean to Middle Palaeolithic transition. In T. Wynn, K. A. Overmann, & F. L. Coolidge (Eds.), *Oxford Handbook of Cognitive Archaeology* (pp. 723–740). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780192895950.013.31

Shipton, C., Clarkson, C., Pal, J. N., Jones, S. C., Roberts, R. G., Harris, C., Gupta, M. C., Ditchfield, P. W., & Petraglia, M. D. (2013). Generativity, hierarchical action and recursion in the technology of the Acheulean to Middle Palaeolithic transition: A Perspective from Patpara, the Son Valley, India. *Journal of Human Evolution*, *65*(2), 93–108. https://doi.org/10.1016/j.jhevol.2013.03.007

Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1050–1059. https://doi.org/10.1098/rstb.2010.0369

Stout, D., & Chaminade, T. (2009). Making tools and making sense: Complex, intentional behaviour in human evolution. *Cambridge Archaeological Journal*, *19*(1), 85–96. https://doi.org/10.1017/S0959774309000055

Wadley, L. (2010). Compound-adhesive manufacture as a behavioral proxy for complex cognition in the Middle Stone Age. *Current Anthropology*, *51*(S1), S111–S119. https://doi.org/10.1086/649836

Seeing is NOT believing: the role of temporal gestures in children's communication with robot and human

Polina Eismont[*1], Anna Zinina[2,3,4], Artemy Kotov[2,3,4], Nikita Arinkin[2,3], Polina Podlesnaya[1], and Maria Yurieva[1]

[*]Corresponding Author: polina272@hotmail.com

[1]Department of General Linguistics named after L. Verbitskaya, St Petersburg University, Saint-Petersburg, Russia

[2]Laboratory of Neurocognitive Technologies, National Research Center "Kurchatov Institute", Moscow, Russia

[3]Laboratory of Anthropomorphic Interfaces, Moscow State Linguistic University, Moscow, Russia

[4]Centre for Cognitive Programs and Technologies, Russian State University for the Humanities, Moscow, Russia

The study focuses on the role of temporal gestures in children's understanding of the order of events in an oral story. A series of experiments with two groups of Russian monolingual children, aged 49 to 73 months, shows that while interpreting the order of events, children rely on temporal conjunctions, iconicity and naturalness. Another significant factor is the distinction between a human and a robot: a human provokes much higher accuracy rates than a robot companion. Temporal gestures played no significant role in children's understanding of the order of events.

The focus of our study is the metaphoric representation of time as space, the development of this cognitive category in children and the assistive role of gestures in its understanding. Time is an abstract concept, metaphorically interpreted as space due to its intangibility (Lakoff, Johnson, 2003). In communication this metaphor can be expressed via temporal gestures that represent time concepts as having spatial extension. In our study, we focus on the evolution of language time/space metaphor in phylogenesis and its acquisition via verbal and nonverbal means in ontogenesis. Our goal is to check whether temporal gestures improve the understanding of the sequence of events in text by Russian native children. Bottini and Casasanto (2013) state that children are able to metaphorically relate time and space by the age of 4, but the formation of this ability continues until the age of 15 (Burns et al., 2019). The role of non-verbal behavior in speech development is, generally, not entirely clear: researchers talk about both the help of gestures in speech acquisition (Hostetter, 2011) and the interference with the use of gestures during speech development (McNeil, Alibali,

Evans, 2000), or even propose a model of parallel development of verbal and non-verbal communication means (Lewis et al., 2000). By the age of 5, children successfully recognize the order of events in speech distinguishing *before* and *after* relations. At the same time, the violation of *naturalness* (where the events have some regular order or cause-and-effect relations) and *iconicity* (where the order of clauses corresponds to the order of real-life events) is the factor that may prevent preschoolers from interpreting the event order correctly (Wagner, Halt, 2023). Basing on the previous studies (Macoun, Sweller 2016; Abner, Cooperrider, Goldin-Meadow 2015) we hypothesize: while non-iconic and non-natural (arbitrary) sentences are difficult for preschoolers, temporal gestures of a tutor compensate for the text complexity and help children correctly recognize the order of events.

To test this hypothesis, we used the emotional robot F-2 (Malkina, Kotov, Zinina 2022) in a series of two experiments. In an experiment the robot can "isolate" the tested communicative cues from other expressive means – thus providing a reliable experimental setup. In the first experiment, we tested the prospects of using temporal gestures by the robot, as an assistive means for preschoolers. The experiment was executed within the program of development of communicative cues for the F-2 companion robot. In the experiment, the robot told children two stories and accompanied them with three types of gestures: (a) left-directed gestures for the preceding action and right-directed gestures for the subsequent action (see Fig. 1); (b) the same groups of gestures but swapped on the horizontal axis, and (c) neutral symmetric gestures, with no pointing to the left or right (control group). Within (a) we assumed that the listener adopts the speaker's metaphor "past is on the left and future – on the right" (Kotov et al., 2024), while in (b) we represented the same metaphor from the point of the participant – so, both spatial representations of time on the transverse axis were tested.

Each story consisted of 4 sentences with two types of constructions: *X before Y* (B) and *Y after X* (A). In addition, the stimuli could be either iconic (I) or non-iconic (N) and either natural (N) or arbitrary (A). For example, *The boy put on his socks before he put on his shoes* is natural (regular order is naturally determined for the events: people put on socks before shoes) and iconic (the order of events in text corresponds to their order in reality; it corresponds to BNI in Fig. 3). Arbitrary non-iconic: *Before the girl jumped, she clapped her hands* (the actions can be performed in any order; the sequence of events is reversed; it corresponds to BAN in Fig. 3). Children were randomly divided in 3 groups; each was presented with stories accompanied by only one of 3 mentioned types of gestures. After each stimulus a child had to name the first of two events. 71 Russian monolingual right-handed children (31 male, mean age 58 months) took

part in the experiment. Each child was interviewed in person. Correct answers were assigned 2 points, incorrect answers – 1 point, and no answer was assigned 0 points. We used ANOVA and T-test for statistical processing of the data. The experiment confirmed that children better understand *before* than *after* sequences (Paired Samples T-Test, $p<0.001$), but the overall results showed a rather low number of correct answers regardless of the condition. The usage of temporal gestures by the robot did not have any significant effect on the number of correct answers.

To test the possible influence of the robot on the acquired results, we compared it with a human tutor in the second experiment. The same stimuli were performed by a tutor and presented as mini-videos (see Fig. 2). 59 Russian monolingual right-handed children (29 male, mean age 61 months) took part in the second experiment. The procedure remained the same. Much higher accuracy rates in the second experiment proved the influence of the type of illocutor on the understanding of event order (ANOVA, $F(8, 126) = 10.592$, $p = 0.001$, see Fig. 3). However, no role of gestures was found either.

Fig. 1. Robot F-2, right-pointing gesture

Fig. 2. Human tutor, a screenshot from a mini-video stimulus

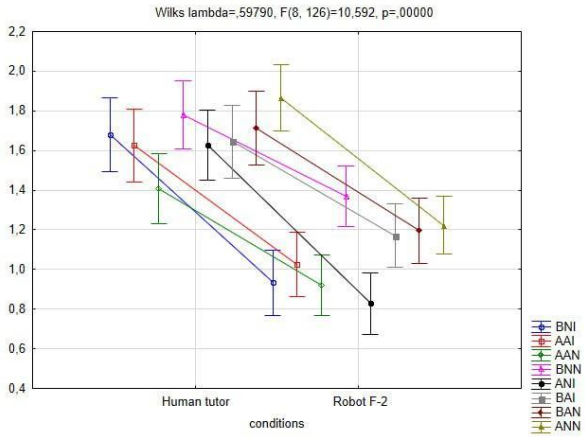


Fig. 3. Accuracy rates in both experiments in all three conditions

We suggest that for children of preschool age the simultaneous processing of temporal conjunctions, possible non-iconic and arbitrary sequences of events

in speech is a complicated cognitive task, thus children tend to block out other communicative input channels, such as non-verbal, relying solely on verbal messages. Choosing a robot as an illocutor only complicates the task.

References

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). *Gesture for linguists: A handy primer*. Language and Linguistics Compass, 9(11), 437–451. https://doi.org/10.1111/lnc3.12168

Bottini R., & Casasanto D (2013) *Space and time in the child's mind: metaphoric or ATOMic?* Front. Psychol. 4:803. doi: 10.3389/fpsyg.2013.00803

Burns, P., McCormack, T., Jaroslawska, A.J., O'Connor, P.A. & Caruso, E.M. (2019), *Time Points: A Gestural Study of the Development of Space–Time Mappings*. Cogn Sci, 43: e12801. https://doi.org/10.1111/cogs.12801

Casasanto, D. & Kyle J. (2012). *The Hands of Time: Temporal Gestures in English Speakers*. Cognitive Linguistics. 23. 10.1515/cog-2012-0020

Hostetter, A. B. (2011). *When do gestures communicate? A meta-analysis.* Psychological Bulletin, 137(2), 297–315.

Kotov, A. A., Zinina, A. A., Arinkin, N. A., & Berkuta, D. V. (2024). Can gestures communicate the time of described events? The perception of temporal gestures executed by a companion robot. *Research Result. Theoretical and Applied Linguistics*, *10*(2), 100–116. https://doi.org/10.18413/2313-8912-2024-10-2-0-5

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. University of Chicago.

Lewis, V., Boucher, J., Lupton, L., & Watson, S. (2000). *Relationships between symbolic play, functional play, verbal and non-verbal ability in young children.* International journal of language & communication disorders, 35(1), 117–127. https://doi.org/10.1080/136828200247287

Macoun, A., & Sweller, N. (2016). *Listening and watching: The effects of observing gesture on preschoolers' narrative comprehension*. Cognitive Development, 40, 68–81. https://doi.org/10.1016/j.cogdev.2016.08.005

Malkina M., Kotov A., & Zinina A. *Perception of Multimodal Hedges Communicative Behavior of a Companion Robot.* In Proceedings of the 10th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). Nara, 2022. P. 1—3.

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). *The role of gesture in children's comprehension of spoken language: Now they need it, now they don't*. Journal of Nonverbal Behavior, 24(2), 131– 150. https://doi.org/10.1023/A:1006657929803

Wagner, L., & Holt, R.F. (2023). *Time after time: Factors influencing children's comprehension of Before and After*. Journal of Child Language 1–9, https://doi.org/10.1017/ S0305000923000612

Sequence representation as limitation on would-be primate speech

Axel Ekström[1,2,3*], Johan Lind[3,4], and Anna Jon-And[3,5]

[1]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[2]Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden
[3]Centre for Cultural Evolution, Dept. of Psychology, Stockholm University, Stockholm, Sweden
[4]Biology Div., Dept. of Physics, Chemistry, and Biology, Linköping University, Linköping, Sweden
[5]Dept. of Romance Studies and Classics, Stockholm University, Stockholm, Sweden
[*]Corresponding Author: axeleks@kth.se

In speech, sequences of speech sounds (or phonemes) are combined into words and sentences. In recent literature, great apes have been shown to possess several vocal capacities previously believed to elude them, including voluntary control of the articulators, and a capacity to produce several speech-like sounds. We identify a core constraint on syllabic speech, downstream of sequence representation capacity, the ability to remember and maintain representations of arbitrary sequences of auditory stimuli.

1. What could primates say?

As early as the 1970s, Lieberman, Crelin, and Klatt (1972, p. 298) argued that the lack of speech in primates "may reflect the absence of required neural mechanisms." It has been known for decades that primates possess the potential for a range of vowel sounds (Lieberman et al., 1972; Fitch, de Boer, Mathur, & Ghazanfar, 2016). Great apes also evidently possess a stark degree of freedom in movements of articulators, facilitating the production of – for example – labial stops (Ekström, Gannon, Edlund, Moran, & Lameira, 2024), clicks (Ekström, 2023) and trills (Lameira, 2017). Apes also possess control over the voice (Lameira & Shumaker, 2019), necessary to produce voiced utterances. Yet great ape vocal repertoires are typically highly limited, bearing little resemblance to speech. A comparison with syllabic speech illustrates a stark difference. Phonemic forms available through four labial consonant sounds (all of which have been reported in apes), with a single unstressed vowel sound, can be expressed as the total number of combinations $> 8^n$, where n is the number of permissible syllables. This results in a minimal number of 24 (4 CV combinations, 4 VC combinations, and $4 \cdot 4$ CVC combinations) total possible syllables. Thus, even incremental phonemic combinatoriality would allow for a dramatic increase in the range of available signals. Yet, there is little to no evidence supporting such combinatoriality in the vocal domain (Girard-Buttoz et al., 2022) in the wild; only exceptionally and in captivity (Fischer & Hammerschmidt, 2020; Ekström et al., 2024).

2. What limits primate speech?

Several ostensible neural limitations have been suggested to answer why apes cannot speak like humans. For example, Brown, Yuan, and Belyk (2021) speculated that, unlike humans, apes did not possess a cortical somatotopic overlap between regions representing the larynx and jaw, precluding "babble"-like utterances – a potential "precursor" to speech in human development. Recent evidence of chimpanzees uttering voiced productions of the phonetic form "mama" (Ekström et al., 2024), shows that either such an overlap exists in apes, or it is not necessary for production of speech-like signals. We argue that instead of inferring the lack of capacities from human experimental data, we may seek to posit specific constraints on linguistic expression. Here, we posit one such constraint relating to the combination of potentially "speech-like" behavior.

3. Sequence representation as prerequisite for speech

Faithful sequence representation refers to the ability to instantly recognize and remember the order of stimuli or events that are experienced from one's surroundings. This ability has been suggested to be uniquely human and a fundamental prerequisite for language, thinking, and cumulative culture on a large scale (Jon-And, Jonsson, Lind, Ghirlanda, & Enquist, 2023). Reviews of animal memory reveal that there is little convincing evidence of faithful sequence representation capacities in non-human animals (Ghirlanda, Lind, & Enquist, 2017; Lind & Jon-And, 2024). Nonhuman animals, from birds to mammals, struggle to distinguish between short sequences of stimuli. For example, they may confuse a red-green sequence with green-red or green-green sequences even after extensive training. Notably, great apes do not perform any better than other animals (Lind, Vinken, Jonsson, Ghirlanda, & Enquist, 2023). In contrast, humans can learn to distinguish such sequences with minimal training, demonstrating a more robust and accurate encoding of sequential information.

Speech sounds occur in rapid succession, and a language requires the ability to process and recall sequences of sounds to understand and produce words accurately. With a reduced memory for sequences, a speaker would struggle with phonological awareness – essential for distinguishing between pairs, such as "tap" and "pat," where the order of sounds changes the meaning. The ability to accurately store and retrieve sequences of speech sounds is fundamental to the comprehension and production of language. Without it, efficient processing and use of language would be markedly impaired. Primates possess the control necessary for executing several speech-like sounds, but may be limited from ritualized use through constraints on sequence representation. If multiple arbitrary sequences of sounds cannot be recognized and represented with precision, a system of speech would likely be out of reach.

Acknowledgements

AE was funded through the Swiss National Science Foundation (PCEFP1_186841).

References

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.

Ekström, A. G. (2023). Viki's first words: A comparative phonetics case study. *International Journal of Primatology*, *44*, 249-253.

Ekström, A. G., Gannon, C., Edlund, J., Moran, S., & Lameira, A. R. (2024). Chimpanzee utterances refute purported missing links for novel vocalizations and syllabic speech. *Scientific Reports*, *14*, 17135.

Fischer, J., & Hammerschmidt, K. (2020). Towards a new taxonomy of primate vocal production learning. *Philosophical Transactions of the Royal Society B*, *375*(1789), 20190045.

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*(12), e1600723.

Ghirlanda, S., Lind, J., & Enquist, M. (2017). Memory for stimulus sequences: a divide between humans and other animals? *Royal Society Open Science*, *4*(6), 161011.

Girard-Buttoz, C., Bortolato, T., Laporte, M., Grampp, M., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2022). Population-specific call order in chimpanzee greeting vocal sequences. *iScience*, *25*(9), 104851.

Jon-And, A., Jonsson, M., Lind, J., Ghirlanda, S., & Enquist, M. (2023). Sequence representation as an early step in the evolution of language. *PLOS Computational Biology*, *19*(12), e1011702.

Lameira, A. R. (2017). Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neuroscience Biobehavioral Reviews*, *83*, 429–439.

Lameira, A. R., & Shumaker, R. W. (2019). Orangutans show active voicing through a membranophone. *Scientific Reports*, *9*(1), 1–6.

Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, neanderthal man, and the chimpanzee. *American Anthropologist*, *74*, 287-307.

Lind, J., & Jon-And, A. (2024). A sequence bottleneck for animal intelligence and language? *Trends in Cognitive Sciences*.

Lind, J., Vinken, V., Jonsson, M., Ghirlanda, S., & Enquist, M. (2023). A test of memory for stimulus sequences in great apes. *Plos One*, *18*(9), e0290546.

Signaling and inferring cooperative urge via speech rhythm

Leona Polyanskaya[*1], Arthur Samuel[2,3], and Mikhail Ordin[4,5]

[*]Corresponding Author: mikhail.ordin@gmail.com
[1]Faculty of Psychology, Universidade de Coimbra, Portugal
[2]Stony Brook University, New York, USA
[3]BCBL, Donostia, Spain
[4]Medical Faculty, Universidade de Coimbra, Portugal
[5]Coimbra Institute for Biomedical Imaging, Portugal

Patterns of speech rhythm in vocalization of animals and in human speech are correlated with the degree of the cooperative urge of the interacting agents and their mutual goodwill. However, there is little evidence whether the third-party observer is able to make pragmatic inferences regarding the cooperative behavior of the interacting agents or their hostility. We set up a series of studies to understand whether humans can map rhythmic patterns in interlocutors' speech onto pro-sociality and cooperative inclinations.

Patterns of speech rhythm in vocalization of animals (Rek & Osiejus, 2013) and in human speech (Pickering & Garrod, 2004) are correlated with the degree of the cooperative urge of the interacting agents and their mutual goodwill (Mehr et al., 2021; Savage et al., 2021). However, there is little evidence whether the third-party observer is able to make pragmatic inferences regarding the cooperative behavior of the interacting agents or their hostility. We set up a series of studies to understand whether humans can map rhythmic patterns in interlocutors' speech onto pro-sociality and cooperative inclinations.

For the analytical purposes, we consider rhythm at two levels: pulse (salient acoustic events, which may recur at regular or irregular intervals) and meter (hierarchical structuring of pulses into groups based on their relative salience). In the first series of experiments, we explored whether regularity in rhythm at the level of pulse (exp.1) and at the level of meter (exp.2) is perceived as a signal of cooperation and social bonding between the interlocutors. Participants – native Spanish speakers – had to listen to a pair of artificial language sentences and they had to report whether the interacting agents are friendly and cooperating with each other, or hostile to each other, and then indicate the confidence in their response at a 4-point scale. In experiment 1, we manipulated vowel durations. Higher variability in duration of vowels leads to lower syllabic isochrony, i.e., less

rhythmic regularity (and predictability when the next syllable starts). In experiment 2, we created 3-syllabic groups, in which either the first or the second syllable was made more prominent by F0 increase and lengthening compared to the other two vowels in the group. Some sentences consisted of rhythmic groups with the same syllable (either the first or the second) that is more prominent than the others, and such sentences were referred to as having regular meter. On other sentences – referred to as sentences with meter irregularities – some syllabic groups had the first, and the other syllabic groups had the second vowel more prominent.

We found that regular pulse is indeed mapped onto the social bonding and cooperative urge, however, regularity at the level of meter is not associated with perceived pro-sociality and cooperation between interlocutors. We suggest that pulse isochrony allows better interpersonal motor and vocal coordination via motor coordination with the acoustic signal emitted by a different individual, which strengthens social bonding and promotes pro-social behavior.

Additionally, we analyzed whether humans are aware of the pragmatic inferences they are making based on patterns of speech rhythm. Analysis of awareness indicators show that participants are more aware that isochrony at the level of pulse signals cooperation and social bonding. Awareness of one's abilities is reflected in a feeling of confidence reported on each particular decision (Maniscalco, B. & Lau, 2012). Higher awareness enables assigning higher confidence ratings to correct responses than to incorrect responses, and consequently assign more or less credit to different information sources (in this case, particular rhythmic patterns), thus calibrating the behavior accordingly. Being aware of one's decisions regarding the mental state (i.e., social bonding and cooperative inclinations within a group under observation) promotes development of intentionality and theory of mind, which is beneficial for the individuals' fitness because the ability to understand the intentions of others allows active manipulation of their behavior and dynamically adjust one's behavior through selective attention to the relevant properties of the communicative signal during language evolution (Dunbar, 2004).

Isochrony in vocalizations may signal cooperative urge by facilitating social entrainment, which is the entrainment of behavior, including verbal behavior, to the signal emitted by a different conspecific individual (Phillips-Silver et al., 2010). In social entrainment, mechanisms of rhythmic cognition and synchronization of the motor output with the input signal are activated by the cues from the social environment and allow coordination of movements and vocalizations, including speech production, and even entrainment of neural oscillations (Bowling et al., 2013; Stephens et al., 2010). This further promotes social bonding (Haidt et al., 2008; Kirschner & Tomasello, 2009), and may signal,

to the third-party observer, that the communicating individuals are socially affiliated. We tested this hypothesis in two further experiments, exploring whether rhythmic entrainment is speech of interacting agents will be perceived as a cue to the cooperative inclination of the the individuals. The results suggest that rhythm convergence can be a marker of social cooperation at the level of pulse, but not at the level of meter. The mapping of rhythmic convergence onto social affiliation or opposition is important at the early stages of language acquisition. The evolutionary origin of this faculty is possibly the need to transmit and perceive coalition information in social groups of human ancestors. The mapping of vocal rhythm convergence onto social affiliation is important for the development of social cognition and for language acquisition in ontogenesis, and probably was an important facilitating factor for speech emergence in phylogenesis.

References

Bowling, D., Herbst, C., & Fitch, T. (2013). Social origin of rhythm? Synchrony and temporal regularity in human vocalization. PLOS One 8(11), e80402.

Dunbar, R. (2004). The human story: a new history of mankind's evolution. Oxford: Faber and Faber.

Haidt, J., Seder, J., & Kesebir, S. (2008). Hive Psychology, Happiness, and Public Policy. Journal of Legal Studies 37, 133-156.

Kirschner, S., & Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. Journal of Experimental Child Psychology 102, 299–314.

Maniscalco, B., & Lau, H. A. (2012). Signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Consciousness and Cognition 21, 422–430.

Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. Behavioral and Brain Sciences, 44, e60. https://doi.org/10.1017/S0140525X20000345

Phillips-Silver,J., Aktipis, A., & Bryant, G. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. Music Perception 28, 3-14.

Pickering, M., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. Bahavioral and Brain Sciences 27, 169-226.

Ręk, P., Osiejuk, T.S. (2013) Temporal patterns of broadcast calls in the corncrake encode information arbitrarily. Behavioral Ecology 24, 547-552.

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, e59. https://doi.org/10.1017/S0140525X20000333

Stephens, G., Silbert, L., Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. PNAS 107, 14425-14430.

Sorries seem to have the harder words

Shiri Lev-Ari*[1]

*Corresponding Author: shiri.lev-ari@rhul.ac.uk
[1]Dept. of Psychology, Royal Holloway, University of London, UK

This paper shows that people use longer words when apologizing (Study 1) and interpret apologies with longer words as more apologetic (Study 2). This supports signaling accounts that propose that apologizers should incur a cost (greater production effort) to indicate their sincerity. This behavior illustrates a new type of iconicity: dynamic iconicity – iconicity that is context-dependent rather than inherent to a word's meaning. These studies have implications for our understanding of the emergence, prevalence, and role of iconicity in communication.

Iconicity in language refers to the non-arbitrary link between form and meaning. Research typically focuses on specific word forms (e.g., /n/ in *nose*) and their meanings (Blasi et al., 2016). This set of studies explores *dynamic iconicity*— the selection of words from the general lexicon whose form conveys intended meaning in context. Specifically, it tests whether people use harder-to-produce words when apologizing to express greater apologeticness.

Apologies are "cheap". Apologizers could emphasize the sincerity of their apology by incurring a cost via use of longer or low frequency words, as these are harder to produce. To test this, 50 apology tweets (25 from celebrities, 25 from regular users) were compared to control tweets by the same users. A mixed effects model revealed that apology tweets consisted of significantly longer words than control tweets (β=0.85, SE=0.17, t=-4.87; See Figure 1). Celebrity status did not influence word length and did not interact with Content Type. Further analyses revealed that the effect was not driven by differences in valence. The effect also remained after removing explicit apology words, indicating that it is not the case that apology words are longer (fixed iconicity) but that individuals select longer words from their general lexicon to signal effort when apologizing. In contrast with word length, apologies and control tweets did not differ in word frequency.

Study 2 tested whether people interpret apologies with longer or less frequent words as more apologetic. Fifty-one participants ranked the relative apologeticness of apology triads consisting of versions using short high

frequency words (e.g., *I did not mean to answer in a hostile way*), short low frequency words (e.g., *I did not mean to reply in a combative style*), and long low frequency words (e.g., *I did not mean to respond in a confrontational manner*). There wasn't a version with long high frequency words as long high frequency words were too rare to find. Apologies were matched in meaning (M=0.94 cosine similarity with the Bert tokenizer in Scikit-learn; Pedregosa et al., 2011), and t-tests confirmed that conditions matched/differed in frequency and length as intended. An intercept-only mixed effects model showed that apologies with long words were ranked as more apologetic than those with short words matched for frequency (β=0.34, SE=0.13, t=2.59), while apologies with short low frequency words were not perceived as more apologetic than those with short high frequency words. That is, in line with Study 1, word length, but not word frequency, influence how apologetic apologies are perceived.

The studies show that incurring a production cost by producing longer words signals and is interpreted as greater apologeticness. In contrast, producing lower frequency words does not have the same effect, potentially because lower frequency words, unlike longer words, are harder to process, so burden the addressee. These studies suggest that iconicity might exist in language not only in the relationship between specific forms and their meaning but also in individuals' dynamic lexical choice. It thus opens a new direction to research on iconicity and its role in communication.

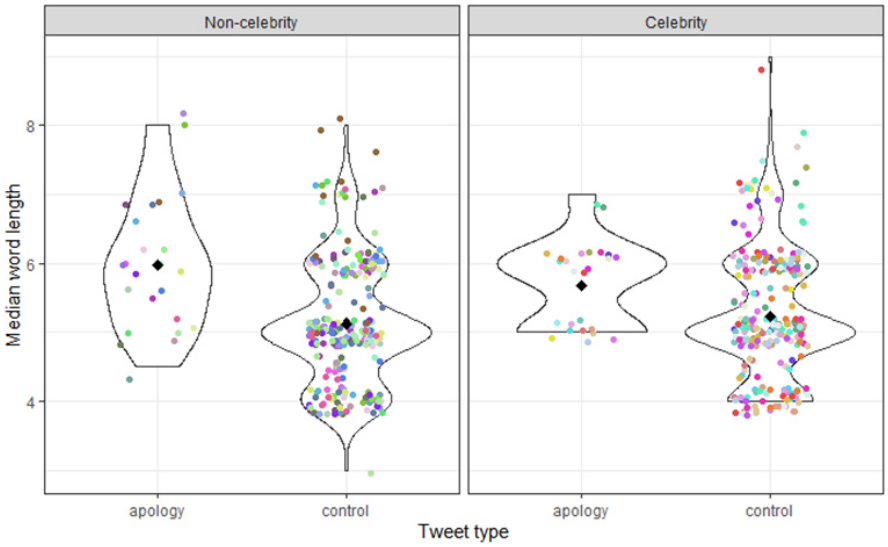


Figure 1. Median word length in tweets by Tweet Type and celebrity status. Each color indicates a different individual. The diamonds indicate condition mean.

References

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *PNAS, 11339*, 10818–10823.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977-990.

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.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.

Rinker, T. W. (2018). sentimentr: Calculate Text Polarity Sentiment version 2.2.3. http://github.com/trinker/sentimentr

Acknowledgements

The research was funded by anonymized funder, grant number.

Statistical Learning of Language from Evolutionary Perspective

Leona Polyanskaya[1,*], Arthur Samuel[2,3], and Mikhail Ordin[4,5]

[*]Corresponding Author: leona.polyanskaya@gmail.com
[1]Faculty of Psychology, Universidade de Coimbra, Portugal
[2]Stony Brook University, New York, USA
[3]BCBL, Donostia, Spain
[4]Medical Faculty, Universidade de Coimbra, Portugal
[5]Coimbra Institute for Biomedical Imaging, Portugal

Although statistical learning mechanisms are recycled for the purposes of language acquisition and speech processing, they are also employed in non-language domains and shared by non-linguistic species. That is why such mechanisms might be more efficient on processing non-linguistic stimuli, for which they initially evolved. We provide an explicit within-subject comparison of the utility of statistical learning in language versus nonlanguage domains across the visual and auditory modalities, and discuss the adaptation of statistical learning for language from evolutionary perspective.

1. Rational of the study

Statistical learning (SL) is a set of neurocognitive mechanisms underlying the ability to extract regularities from the environment: recurrent patterns and sequences, transitional probabilities (TPs), probabilities that one event predicts a subsequent event (transitional probabilities – TPs – between the events or recurrent patterns). The neurocognitive mechanisms underlying SL are engaged when humans listen to natural speech, and performance on SL tasks is correlated with linguistic abilities (Erickson& Thiessen, 2015). However, SL also operates on nonlinguistic material (Gebhart et al., 2009) and has been observed in a range of taxonomically different species that do not have a language faculty (Kikuchi et al., 2018; Milne et al., 2018). We suggest that SL mechanisms are evolutionarily ancient, making it highly unlikely that they evolved specifically to process linguistic input, and hence they might be more effective for processing environmental stimuli than linguistic stimuli. We tested this hypothesis in an explicit within-subject comparison of the utility of statistical learning in language versus nonlanguage domains across the visual and auditory modalities.

2. Experimental Design

We used a Saffran-style paradigm to explore SL in a 2*2 design (2 modalities – visual and auditory – and 2 domains – linguistic and non-linguistic). For linguistic

material, we created a continuous stream of 32 syllables, 24 were arranged into recurrent triplets with TPs between syllable within triplet is 1.0, and the other 8 syllables used as fillers between triplets (following Gervain et al., 2008), to model function words and to create a stream more language-like. The TPs between fillers and triplet-initial or triplet-final syllable was 0.125. Additionally, intonational contour was imposed to create prosodic frames for phrases and utterances (utterance is made up of 2 phrases). Non-linguistic stimuli included environmental sounds (footsteps, waterdrops, creaks, animal cries) following the same statistical structure (ramping was used for hierarchical structuring instead of intonation, which made the stimuli as rich in prosodic cues as linguistic stimuli, although prosodic cues were not those typically employed in natural speech).

For the visual linguistic stimuli, we used syllables (set of syllables was different from that used in the auditory modality), and for non-linguistic stimuli we created fractals: units were concatenated following the same statistical structure as the auditory stimuli, with punctuation marks (linguistic) or squares (non-linguistic) used for hierarchical structuring in the same positions where boundary tones or ramping occur in the auditory modality. Visually, the syllables or fractals were presented one by one in the middle of the screen for the familiarization. Following the familiarization, people had to do the test, when they listened/saw a triplet or a foil composed of the same elements and had to report whether it was a "word" from the alien language they were exposed to or not.

3. Results and Interpretations

The results clearly showed that SL in the visual modality was significantly better (when measured as a d') on non-linguistic than on linguistic domain, while in the auditory modality statistical learning was better in linguistic than in non-linguistic domain. These results were further replicated in three other groups (4 groups in total: Basque bilinguals, Catalan bilinguals, Spanish monolinguals, and mixed group), with two different sets of experimental materials. We suggest that the speech faculty has been important for individual fitness for an extended period, leading to the adaptation of statistical learning mechanisms for speech processing. This is not the case in the visual modality, in which linguistic material presents a less ecological type of sensory input. SL was shaped for processing nonlinguistic environmental stimuli and only later, as the language faculty emerged, recycled for speech processing. This led to further adaptive changes in the neurocognitive mechanisms underlying speech processing, including SL. By contrast, as a recent cultural innovation, written language has not yet led to adaptations.

Further inter-group analysis showed that auditory SL can be further modulated by exposure to a bilingual environment, in which speakers need to process a wider range of diverse speech cues. This effect was observed only in language domain. We conclude that ontogenetic factors modulate the efficiency of already existing

SL ability, honing it for specific types of input, by providing new targets for selection via exposure to different cues in the sensory input.

References

Erickson, L.C. & E.D. Thiessen. 2015. Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66–108.

Gebhart, A.L., E.L. Newport & R.N. Aslin. 2009. Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin and Review 16*, 486–490.

Gervain, J., M. Nespor, R. Mazuka, et al. 2008. Bootstrapping word order in prelexical infants: a Japanese–Italian cross-linguistic study. *Cognitive Psychology, 57*, 56–74

Kikuchi, Y., W. Sedley, T.D. Griffiths & C.I. Petkov. 2018. Evolutionarily conserved neural signatures involved in sequencing predictions and their relevance for language. *Current Opinions in Behavioral Sciences, 21*, 145–153.

Milne, A.E., C.I. Petkov & B. Wilson. 2018. Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Journal of neuroscience, 389*, 104–117.

Syllable sequence in adult pale spear-nosed bats (*Phyllostomus discolor*)

Newar SL[*1], Jenks T[1], Moran IG[2], Crighton AC[1], and Vernes SC[1]

[*]Corresponding Author: sln5@st-andrews.ac.uk
[1]School of Biology, University of St Andrews, St Andrews, United Kingdom
[2]Department of Integrative Biology, University of Windsor, Windsor, Canada

Vocal syntax is a complex form of communication in which behavioural information is communicated around precise syllable sequences. The genetic and cultural bases of vocal syntax in non-human animals have been poorly studied and it remains unclear whether communication follows syntactical rules in most animals, including bats. We investigated and found that syllable sequence in a group of pale spear-nosed bats follows predictable syntactical rules, with these sequences being disrupted in deafened bats.

1. Introduction

To grasp the origins of human language in the absence of fossil evidence, we can turn to studying the evolution of language-relevant traits like vocal communication in non-human animals. Syntax, the production of multiple call types or words organized in predictable combinations depending on various behavioural or social contexts, is a key trait in human language that has also been demonstrated in many animals (Zhang, *et al.*, 2019). While it has been argued that syntactical rules in animals are not equivalent to the complex and hierarchical syntax in humans (Jaber, Omari, & Abudalbuh, 2021), the presence of predictable call sequences and the use of calls in multiple combinations in animals could inform how our unique communication systems first originated.

2.1. Call syntax in bats

Vocalizations in bats are essential for communication between individuals. With the potential to encode internal states and facilitate complex social interactions – acoustic communication is vital for bats who have rich social structures and are most active in dark environments with little visual information (Kanwal *et al.,* 1994). Call syntax has been demonstrated in multiple species of bats in the context

of agonistic interactions (*Rhinolophus ferrumequinum*; Zhang, *et al.,* 2019; *Rousettus aegyptiacus*; Amit & Yovel, 2023), singing (*Tadarida brasiliensis*; Bohn, Smarsh, & Smotherman, 2013), and group interactions (*Pteronotus parnellii*; Kanwal *et al.* 1994). These species have been shown to produce call sequences in varied contexts (Amit & Yovel, 2023; Bohn, Smarsh, & Smotherman, 2013; Zhang, *et al.,* 2019), sequence lengths, and syllable combinations (Amit & Yovel, 2023; Bohn, Smarsh, & Smotherman, 2013; Kanwal *et al.* 1994; Zhang, *et al.,* 2019). However, it was also shown in Egyptian fruit bats that while increased sequence length was more likely to result in an accurate response from conspecifics (e.g. moving away when the conspecific is too close to the caller), the syllable order within the call didn't affect the accuracy of the response (Amit & Yovel, 2023), suggesting that syntactical rules aren't necessary for information dissemination in this species.

2.1. Pale spear-nosed bats

Pale spear-nosed bats (*Phyllostomus discolor*) are highly gregarious bats which have been shown to produce at least 13 distinct call classes in captive populations with some calls produced in combination (Lattenkamp *et al.*, 2019). These bats have previously been demonstrated as excellent candidate models for studying vocal learning (Lattenkamp, Vernes, & Wiegrebe, 2020; Vernes *et al.,* 2022), with this ability being described in only a handful of other mammals (Janik & Slater, 1997), and no others being as practical to house in large, long-term captive colonies.

To investigate vocal learning in these bats, we examined if these call combinations follow predictable syntactical rules as shown in other bat species. Similar to other syntactical work done in bats (Kanwal *et al.,* 1994), we used recordings from captive *Phyllostomus discolor* in small groups to facilitate naturalistic vocalizations. We compared groups of adult bats deafened at 9-11 days old and control hearing bats of similar age (Lattenkamp *et al.,* 2021). Vocalizations were classified by multiple independent observers and syllable combinations were analysed using *DeepSqueak* (Coffey, Marx, & Neumaier, 2019). We found in hearing bats that several call types are repeated in monosyllabic sequences while others were produced in combination. Additionally, we found that syllable combinations changed with deafened bats suggesting that the development of syllable sequence requires auditory input.

Using these data, we are exploring call sequence in a group context, where individual call identity was not derived due to the proximity of the callers, and

consequently syntax of an individual caller was difficult to determine. However, we were still able to derive predictable syntactical rules over multiple call types, implying that syllable combinations are important for vocal communication in the context of either single or multiple callers. Additionally, the differences found between deafened and hearing groups implies that auditory input is required to produce the correct sequence of syllables, further supporting that *P. discolor* is a vocal learning species. This work provides a foundation to further investigate call sequence in different caller combinations, age and sex groups, behavioural conditions, under genetic manipulation, and in the context of vocal development and learning.

Acknowledgements

We would like to acknowledge that this work would not be possible without the methodological design and data collection from Lutz Wiegrebe, Ella Z. Lattenkamp, and Meike Linnenschmidt.

References

Amit, Y., & Yovel, Y. Bat vocal sequences enhance contextual information independently of syllable order. *iScience, 26,* 106466.

Bohn, K. M., Smarsh, G. C., & Smotherman, M. (2013). Social context evokes rapid changes in bat song syntax. *Animal Behaviour, 85,* 1485-1491.

Coffey, K. R., Marx, R. E., & Neumaier J. F. (2019). DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations.

Jaber, A., Omari, O., & Abudalbuh, M. (2021). Reflections on protolanguage: evidence from studies on animal communication. *Journal of Educations and Social Research, 11,* 113-123.

Janik, V. M., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behaviour, 26,* 59-99.

Kanwal, J. S., Matsumura, S., Ohlemiller, K., & Suga, N. (1994). Analysis of acoustic elements and syntax in communication sounds emitted by mustached bats. *Journal of the Acoustical Society of America, 96,* 1229-1254.

Lattenkamp, E. Z., Linnenschmidt, M., Mardus, E., Vernes, S. C., Wiegrebe, L., Schutte, M. (2021). The vocal development of the pale spear-nosed bat is dependent on auditory feedback. *Philosophical Transactions of Royal Society B, 376,* 20200253.

Lattenkamp, E. Z., Shields, S. M., Schutte, M., Richter, J., Linnenschmidt, M., Vernes, S. C., & Wiegrebe, L. (2019). The vocal repertoire of pale spear-nosed bats in a social roosting context. *Frontiers in Ecology and Evolution*, *7*, 116.

Lattenkamp, E. Z., Vernes, S. C., & Weigrebe, L. (2020). Vocal production learning in the pale spear-nosed bat, *Phyllostomus discolor. Biology Letters, 16,* 20190928.

Vernes, S. C., Devanna, P., Hörpel, S. G., Alvarez van Tussenbroek, I., Firzlaff, U., Hagoort, P., Hiller, M., Hoeksema, N., Hughes, G. M., Lavrichenko, K., Mengede, J., Morales, A. E., & Wiesmann, M. (2022). The pale spear-nosed bat: A neuromolecular and transgenic model for vocal learning. *Annals of the New York Academy of Sciences, 1517,* 125-142.

Zhang, K., Liu, T., Liu, M., Li, A., Xiao, Y., Metzner, W., & Liu, Y. (2019). Comparing context-depending call sequences employing machine learning methods: an indications of syntactic structure of greater horseshow bats. *Journal of Experimental Biology, 222*, jeb214072.

The Evolutionary Costs of Sequence Representation: Why Only Humans Developed Language

Anna Jon-And[1,2*], Markus Jonsson[1,3], Johan Lind[1,4], and Magnus Enquist[1]

[1]Centre for Cultural Evolution, Deparment of Psychology, Stockholm University, Stockholm, Sweden
[2]Department of Romance Languages and Classics, Stockholm University, Stockholm, Sweden
[3]Karolinska Institutet, Science for Life Laboratory, Stockholm, Sweden
[4]Department of Physics, Chemistry and Biology, Linköping University, Linköping, Sweden
[*]Corresponding Author: anna.jon-and@su.se

Human language is uniquely compositional, sequential, and open-ended. Traditional theories of language evolution focus on the communicative advantages of language, but struggle to explain why language has evolved only in humans and not in other species. This paper proposes that sequence representation, the ability to recognize, store, and recall ordered sequences of information, constitutes a crucial early step in the evolution of language. By integrating formal evolutionary analyses and computational simulations, we demonstrate that while sequence representation is highly costly in terms of memory and learning, it is a fundamental prerequisite for the emergence of language and other cognitive abilities unique to humans.

1. Sequential Abilities in Animals and Humans

A defining feature of human language is compositionality, the ability to build complex expressions from smaller units(Szabó, 2012). Humans are able to insert unknown linguistic elements in known structures (Berko, 1958). This open-endedness of linguistic compositionality would not be possible without faithful sequence representation. Although certain animals exhibit rudimentary forms of combinatorial communication(Suzuki, Wheatcroft, & Griesser, 2017; Zuberbühler, 2002; Leroux et al., 2023), these lack the open-ended and productive compositional structure seen in human languages(Townsend, Engesser, Stoll, Zuberbühler, & Bickel, 2018). Studies suggest that non-human animals recognize sequences approximately rather than faithfully, relying on mechanisms like trace memory, which does not retain precise order information(Ghirlanda, Lind, & Enquist, 2017; Enquist, Ghirlanda, & Lind, 2023; Lind, Vinken, Jonsson, Ghirlanda, & Enquist, 2023). This limitation might explain why language, which requires an accurate representation of sequential order, has not emerged in other species.

2. The Costs and Evolutionary Constraints of Sequence Representation

Mathematical modeling and simulations reveal that learning and retaining sequential information is highly costly due to the combinatorial explosion of possible sequences that an organism must learn to recognize and respond to. Recognizing and responding to single stimuli is significantly more efficient in most environments. Sequence representation is beneficial only when the environment contains an abundance of sequentially structured information and when organisms have prolonged learning opportunities, conditions that were likely met during human prehistory. The sequence hypothesis postulates that these conditions need to reach a critical limit in order for sequence representation to evolve, a limit so unlikely to reach that it has only happened once. Once this threshold is passed, it allows for language and other cumulative culture to emerge over generations, creating a selective advantage for even longer learning times. This co-evolutionary scenario is compatible with the unusually long childhood of humans.

3. Approximate vs. Accurate Sequence Representation

While a trace memory strategy is effective in most nature-like environments, it fails in environments where sequence order carries critical meaning. We introduce the concept of "Flexible Sequence Representation," which implies the ability to flexibly represent and respond to all subsequences within a sequence of a given length. This allows for retaining sequential information efficiently while filtering out irrelevant information. Learning simulations demonstrate that this representation significantly reduces learning costs compared to more rigid sequence storage methods, making it a plausible early evolutionary step towards language (Jon-And, Jonsson, Lind, Ghirlanda, & Enquist, 2023). Such a flexible sequence representation mechanism can generate parsimonious hierarchical representations when processing linguistic input (Jon-And, Michaud, et al., 2020; Jon-And & Michaud, 2024), and combined with generalization it enables the emergence of rudimentary grammatical categories, necessary for productive compositionality (Jon-And, Michaud, et al., 2024).

The findings suggest that the evolution of accurate sequence representation provided a minimal foundation for uniquely human cognitive abilities, including complex language, planning, and cumulative culture (Enquist et al., 2023; Lind & Jon-And, 2024). This aligns with theories that give cultural evolution and learning a central role in the emergence of these abilities (Heyes, 2018; Kirby, Cornish, & Smith, 2008). This paper offers a novel perspective on the evolution of language, arguing that the development of sequence representation was a critical first step. By overcoming the inherent costs of sequence learning through an unusually long childhood and flexible representation strategies, early humans may have unlocked the cognitive capacities necessary for language and cultural transmission. These

insights bridge the gap between animal cognition and human linguistic abilities, providing a plausible trajectory for the evolution of language and higher-order thinking.

References

Berko, J. (1958). The child's learning of english morphology. *Word*, *14*(2-3), 150–177.

Enquist, M., Ghirlanda, S., & Lind, J. (2023). *The human evolutionary transition: From animal intelligence to culture*. Princeton University Press.

Ghirlanda, S., Lind, J., & Enquist, M. (2017). Memory for stimulus sequences: a divide between humans and other animals? *Open Science*, *4*(6), 161011.

Heyes, C. (2018). *Cognitive gadgets: the cultural evolution of thinking*. Harvard University Press.

Jon-And, A., Jonsson, M., Lind, J., Ghirlanda, S., & Enquist, M. (2023). Sequence representation as an early step in the evolution of language. *PLOS Computational Biology*, *19*(12), e1011702.

Jon-And, A., & Michaud, J. (2024). Usage-based grammar induction from minimal cognitive principles. *Computational Linguistics*, *50*(4), 1375–1414.

Jon-And, A., Michaud, J., et al. (2020). Minimal prerequisits for processing language structure: A model based on chunking and sequence memory. In *Evolang xiii, 14-17 april 2020, brussels, belgium* (p. 200-209).

Jon-And, A., Michaud, J., et al. (2024). Emergent grammar from a minimal cognitive architecture. In *Evolang xv.*

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

Leroux, M., Schel, A. M., Wilke, C., Chandia, B., Zuberbühler, K., Slocombe, K. E., & Townsend, S. W. (2023). Call combinations and compositional processing in wild chimpanzees. *Nature Communications*, *14*(1), 2225.

Lind, J., & Jon-And, A. (2024). A sequence bottleneck for animal intelligence and language? *Trends in Cognitive Sciences*.

Lind, J., Vinken, V., Jonsson, M., Ghirlanda, S., & Enquist, M. (2023). A test of memory for stimulus sequences in great apes. *Plos one*, *18*(9), e0290546.

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2017). Wild birds use an ordering rule to decode novel call sequences. *Current Biology*, *27*(15), 2331–2336.

Szabó, Z. (2012). The case for compositionality. *The Oxford handbook of compositionality*, *64*, 80.

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

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal behaviour*, *63*(2), 293–299.

The rise and fall of center embedding: Evidence from non-linguistic sequences in art and iterated learning

Emily Davis*[1], Robert Kluender[2]

*Corresponding Author: emily.d.avian.47@gmail.com
[1]Center for Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland
[2]Department of Linguistics, University of California, San Diego

We suggest that linguistic recursion is an exaptation from nonlinguistic cognitive faculties. Recursive center-embedded patterns appear frequently in ritual-based cultural practices: music, narrative, visual arts. We present the results of a small Twitter-based corpus study showing the predominance of emoji strings with center-embedded structure, and an iterated-learning study showing attrition of center-embedded patterns in similar sequences. These results suggest that center-embedding may emerge, but not necessarily persist, in transmitted human culture, a pattern with possible parallels in diachronic language change.

1. Introduction

The question of how linguistic recursion originated has long been contentious. Hauser et al. (2002) raised the possibility that it was exapted from pre-existing cognitive faculties, such as numeracy, navigation, or social cognition. Here we discuss a possible origin in ritualistic behavior. Recursive center embedding—defined as nested non-adjacent dependencies of matched elements, as in pairs of brackets ({ [ ] }, { [< >] }) or palindromes (ABA, ABCBA, ABACABA), appears in several domains of human activity and artistic practice linked to ritual. Center embedding requires recursive rules in both linguistic and non-linguistic sequences (Chomsky, 1957; Pullum & Rogers, 2006; Rohrmeier et al., 2015) and figures prominently in experimental studies of recursive pattern recognition/generation in both animals and LLMs (Ferrigno et al., 2020; Liao et al., 2022; Hao et al., 2024). We present evidence that center-embedded strings of icons appear naturally in social media usage, yet seem to disappear in natural and experimental iterated transmission. These opposite tendencies suggest a possible evolutionary scenario for the rise and fall of center embedding in the evolution of linguistic recursion.

2. Recursive patterns outside language: Rituals and art

In a series of papers on Vedic ritual, Indologist Frits Staal (1979, 1980, 1984a&b, 1990) identified patterns of embedding strikingly parallel to linguistic recursion: performances of one ritual can be embedded in another (Staal, 1979: 16) like the constituents of a sentence. A ritual may also be preceded or followed by pairs of activities that correspond to each other in some way (Staal, 1980:133). Staal proposed that center-embedding in ritual was exapted into language, giving rise to recursive structure. Comparable patterns occur in some non-human primate and avian behaviors, a possibility suggested by Staal (1984:410) and confirmed by quantitative evidence (Sainburg et al., 2019; Lameira et al., 2024), suggesting that these abilities are not unique to humans.

Recursive patterns appear in other domains of human culture, including music (Hofstadter, 1979; Rohrmeier et al., 2015), narratives (Van Otterlo, 1944; Welch, 1981; Forte & Smith, 2014), and dance choreography (see Kluender & Davis, 2020 and Kluender et al., 2022 for further discussion). In western classical music, for example, the sonata form follows a palindromic ABA pattern (the development, elaboration, and recapitulation of a theme), and there are still more elaborate structures such as ABACABA and the complex ternary form ABA-CDC-ABA (DeVoto, 2017; Naylor, n.d.). Recursive forms also occur in epic narratives. The nested frame story structure of Indian epics like the *Mahābhārata*, in which stories are repeatedly embedded one within the other prior to their completion, has been shown to reflect the structure of the rituals that serve as the narrative framework for the recounting of those tales (Minkowski, 1989). These embedded patterns may also serve to elucidate and justify ritual structures (Witzel, 1987). Other narratives derived from Indian culture such as *The Arabian Nights* exhibit a similar multiply center-embedded frame structure (Irwin, 1994). Another type of recursive pattern in literature involves scenes that form parallel prologues and epilogues to a central event, as found in many classical epics, such as the *Iliad* and the *Odyssey* (van Otterlo, 1944; Whitman, 1958), and medieval works like *Beowulf* and *La Chanson de Roland* (Niles, 1973, 1979). Music and narrative are both historically and culturally linked to ritual (Minkowski, 1989; Merker, 2009), suggesting a common origin for these symmetric patterns.

These recursive symmetrical patterns in literary, musical, and performance arts also have parallels in the visual arts, in the form of bilateral symmetry. For example, as discussed by Mackay et al. (1999), the narrative ring structure in the Homeric epics are analogous to the symmetry and nested frames of the (contemporary) Geometric style of Greek art. Likewise, mirrored symmetry is an important principle in architecture, ranging from the ancient Indian school of *vāstu śāstra* with its close links to Vedic ritual (Chakrabarti, 2013; Meister, 1983), to many modern examples. Like recursive cognition, symmetrical construction may not be unique to humans: for the purpose of courtship rituals, bowerbirds construct symmetric stages (Keagy, 2021) and pufferfish create sand circles (Matsuura, 2015).

Symmetrical center-embedded strings of emoji or other special characters also appear in modern-day internet popular culture. These may be used to bracket words, adding emphasis or tone (McCulloch, 2020:127), or to supplement or illustrate text (McDonald, 2024). Interestingly (especially in light of Staal's observations of ritual "bracketing"), such nested emoji sequences also play a role in contemporary online occultism, among Tumblr and Twitter neo-pagan communities, in the form of "emoji spells." In this practice, a string of emoji representing the caster's intention is bracketed on either side by particular emoji such as a crystal ball or a star, or the whole string of emoji may be palindromic (Towers, 2015; Duca, 2016).

To examine the occurrence of recursive patterns in emoji usage on social media quantitatively, we searched a publicly available Twitter corpus (Ma, 2017) containing ~5 million tweets from 2016-2017, and extracted those containing multiple emojis. These strings were in turn classified into patterns

including palindromes, "brackets," and repeat substrings. 2701 of 22,560 multi-emoji strings (12%) contained a palindromic or bracketed pattern (1-2), 2035 (9%) contained an A^nB^n pattern (3), and 981 (4%) showed a repeating $(AB)^n$ pattern (4), suggesting a preference for mirror symmetry over simple repetition.

(1) 🌙💤🌟💛🌟💤🌙

(2) 🐦‍⬛🐤🐥my darlings🐥🐤🐦‍⬛

(3) 😍😍😍💖💖💖

(4) 🔥👀🔥👀🔥👀🔥👀

3. The emergence and attrition of center-embedding: Observational and experimental evidence

While prevalent in nonlinguistic contexts, center embedding presents processing and working-memory challenges for language, in which semantic reference and time pressure impose greater processing demands. Compared to branching patterns, center embedding is particularly difficult to learn in meaningful artificial languages (Davis & Smith, 2023). Center-embedded non-semantic sequences (e.g. nonsense syllables or letters) are likewise difficult to learn (Öttl et al., 2015; Udden et al., 2012; Moreton et al., 2021).

Similar considerations apply to center-embedded emoji sequences on social media. Informal observations of group text threads in which we were included suggested that palindromic emoji sequences accompanying written text messages emerge spontaneously under such circumstances. These strings were sometimes picked up in modified form by other members of the group in replies, but otherwise tended to disappear and revert to simpler sequences, as shown in sample thread sequences of emojis stripped of accompanying text in Figure 1.

Figure 1: Evolution of center-embedded emoji strings in text conversations.

We further investigated both of these phenomena, the emergence and the disappearance of center-embedded sequences, in an iterated learning paradigm with strings of nonlinguistic icons. We hypothesized that the occurrence of a

repeated icon in the input strings (as opposed to a string composed entirely of different icons) would lead to the formation of further center-embedded patterns, like those which occur in emoji sequences. This hypothesis was confirmed, but embedded strings did not occur at above chance rates. A follow-up study showed that if initial input strings contain center embedding, it disappears in transmission, and is not stable if it re-emerges (Davis, 2023b).

These findings may have parallels in diachronic language change: multiple center-embedding is more prevalent and tolerated in languages with verb-final word order (Davis, 2023a), and there is a general tendency for languages to shift from verb-final to verb-initial order (Givon, 1979, Gell-Mann & Ruhlen, 2011) , a change that may be motivated in part by avoidance of center embedding (Ogura, 2001, 2004; Lorido, 2022).

We conclude that center-embedding has deep roots in human—and possibly nonhuman—culture and cognition, with examples spanning space and time from millennia-old ritual traditions to modern social media. From these faculties, recursion may have been exapted to language in the form of center-embedding. However, compared to the ubiquity of center-embedding in art and performance –from rituals to emoji strings—it is relatively restricted in language. In physical media such as sculpture, pottery, and painting, embedded sequences are fixed, and thus not subject to time or memory constraints. Performance art (ritual, narrative, music, dance) unfolds in real time, but follows an established template dependent only on long-term memory. In temporary art forms like communicative emoji, such sequences do face both time and working memory constraints, and therefore tend to be shorter-lived. Yet by virtue of the fact that they exist in a visuospatial medium, emoji sequences still allow serial search for matched pairs when center-embedded. The processing pressures imposed by the more ephemeral nature of real-time language use, whether signed or spoken, afford no such luxury: there is no external cultural artifact available to facilitate serial search independent of working memory. This is probably the main reason that center-embedding in language tends to yield in favor of branching patterns over time. The working memory burden of processing $subject_1$-$subject_2$-$predicate_2$-$predicate_1$ sequences can easily be avoided by simple repetition of concatenated $subject_1$-$predicate_1$ $subject_2$-$predicate_2$ sequences—likely facilitated by processes of cultural transmission. We thus speculate that on a longer time scale, linguistic recursion may likewise have started out as center-embedding before largely shifting to the use of branching structures instead.

Acknowledgements

Emily Davis was supported by the National Science Centre of Poland (NCN) under grant agreement UMO-2024/53/B/HS2/01366.

References

Chakrabarti, V (2013). *Indian architectural theory and practice: Contemporary uses of Vastu Vidya.* Routledge.

Chomsky, N. (1957). *Syntactic structures*. Mouton & Co.

Davis, E. (2023a). Multiple center-embedding is more common in verb-final languages. Presented at LSA 2023.
Davis, E. (2023b). The nature and evolution of center-embedding: an interdisciplinary approach. PhD dissertation.
Davis, E., & Smith, K. (2023). The learnability and emergence of dependency structures in an artificial language. *Journal of Language Evolution*, lzad006.
DeVoto, M. (2017) Ternary form. In Encyclopedia Britannica. Retrieved from https://www.britannica.com/art/ternary-form. Accessed 2025-05-23.
Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, US adults, and native Amazonians. *Science Advances, 6*(26), eaaz1002.
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.
Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences, 108*(42), 17290-17295.
Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*(7088), 1204-1207.
Givón, T. (1979) *On Understanding Grammar.* Academic, New York.
Hao, Y., Angluin, D., & Frank, R. (2022). Formal language recognition by hard attention transformers: Perspectives from circuit complexity. *Transactions of the Association for Computational Linguistics, 10*, 800-810.
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.
Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An Eternal Golden Braid.* Basic Books.
Irwin, R. (1994). *The Arabian nights: A companion.* London: Penguin Books.
Keagy, J. (2021). Bowerbird Innovation and Problem-Solving. In A. Kaufman, J. Call, & J. Kaufman (Eds.), *The Cambridge Handbook of Animal Cognition* (pp. 667-689). Cambridge: Cambridge University Press.
Kluender, R. & Davis, E. (2020). The role of ritual in the evolution of language and the emergence of recursion. EvoLang conference 2020 (canceled due to COVID).
Kluender, R., Davis, E., Rolnik, M. (2022). The recursion those who study language seek is everywhere if you know how to look for it. Poster presented at JCOLE 2022.
Lameira, A. R., Hardus, M. E., Ravignani, A., Raimondi, T., & Gamba, M. (2024) Recursive self-embedded vocal motifs in wild orangutans. *eLife 12.*
Liao, D. A., Brecht, K. F., Johnston, M., & Nieder, A. (2022). Recursive sequence generation in crows. S*cience Advances, 8*(44), eabq3356.
Lorido, R. P. (2022). The role of (the avoidance of) centre embedding in the change from OV to VO in English. In Los, B. et al. (eds.), *English Historical Linguistics: Change in Structure and Meaning. Papers from the XXth ICEHL, 137.*
Ma, M. (2017). Chat Corpus. https://github.com/Marsan-Ma-zz/chat_corpus

Mackay, A., Harrison, D., & Masters, S. (1999). The Bystander At The Ringside: Ring-Composition In Early Greek Poetry And Athenian Black-Figure Vase-Painting. In Signs of Orality: The Oral Tradition and Its Influence in the Greek and Roman World. (pp. 115-142). Brill.

Matsuura, K. (2015). A new pufferfish of the genus Torquigener that builds “mystery circles” on sandy bottoms in the Ryukyu Islands, Japan (Actinopterygii: Tetraodontiformes: Tetraodontidae). *Ichthyol. Res. 62*, 207–212. https://doi.org/10.1007/s10228-014-0428-5

McCulloch, G. (2020). *Because internet: Understanding the new rules of language.* Penguin.

McDonald, L. (2024). Making sense of emoji. *Philosophy, 99*(3), 413-435.

Meister, M. (1983). *Encyclopaedia of Indian temple architecture*. University of Pennsylvania Press.

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.

Moreton, E., Prickett, B., Pertsova, K., Fennell, J., Pater, J., & Sanders, L. (2021). Learning repetition, but not syllable reversal. In *Proceedings of the Annual Meetings on Phonology* (Vol. 8)

Naylor, M. (n.d.). Abacaba.org https://abacaba.org/ Accessed 2025-05-23.

Niles, J. D. (1973). Ring-Composition in "La Chanson de Roland" and "La Chançun de Willame". *Olifant,* 1(2), 4-12

Niles, John D. (1979). Ring Composition and the Structure of Beowulf. *PMLA.* 94 (5): 924–35.

Ogura, M. (2001). Perceptual factors and word order change in English. *Folia linguistica historica, 35*(1-2), 233-254.

Ogura, M. (2004). Evolution of word order. *Folia linguistica historica, 38*(1-2), 21-40.

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.

Pullum, G. K., & Rogers, J. (2006). Animal pattern-learning experiments: Some mathematical background. Ms. Radcliffe Institute for Advanced Study/ Harvard University.

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 B: Biological sciences*, 370.

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

Staal, F. (1979). The meaninglessness of ritual. *Numen,* 26(1), 2-22.

Staal, F. (1980). Ritual syntax. In Nagatomi, M., Matilal, B. K., Masson, J. M., & Dimock, E. C. (Eds.), Sanskrit and Indian Studies (pp. 119-142). Springer Netherlands.

Staal, F. (1984a). Ritual, mantras and the origin of language. In Joshi, S. D. (Ed.) Amrtadhara: Professor R. N. Dandekar Felicitation Volume, (pp. 403-425). Ajanta Publications

Staal, F. (1984b). The search for meaning: music, mathematics, and ritual. The *American journal of semiotics, 2*(4), 1-57.
Staal, F. (1990). R*itual and mantras: Rules without meaning.* Motilal Banarsidass Publications.
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
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.
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 OldIndian literature. In H. Falk (Ed.), H*induismus und Buddhismus: Festschrift für Ulrich Schneider* (pp. 380-414). Freiburg: Hedwig Falk Verlag.
Wright, C. (2016). *Listening to Western music.* Cengage Learning.

The Roots of Facial Expressions: Pragmatic and Semantic Facial Expressions Comprise Biologically-Rooted Facial Movements

Jonas Nölle[*1,2], Yichen Wu[2], Pablo Arias Sarah[2,3], Chaona Chen[2,4], Oliver G.B. Garrod[2], Philippe G. Schyns[2], and Rachael E. Jack[2]

[*]Corresponding Author: jonas.noelle@glasgow.ac.uk
[1]Centre for Language Evolution Studies,
Nicolaus Copernicus University in Toruń, Poland
[2]School of Psychology & Neuroscience, University of Glasgow, UK
[3]Department of Cognitive Science, Lund University, Sweden
[4]School of Computer Science, University of Sheffield, UK

We explore the biological roots of facial expressions, suggesting they evolved from sensory-regulating movements. Analyzing Western and East Asian participants, we find that expansive facial movements (e.g., raised eyebrows) consistently convey high arousal and positive messages, while contractive movements (e.g., squinting) indicate negative messages. These cues also impact speech perception, with expansion movements signaling confidence and larger quantities, and contraction movements signaling doubt and smaller amounts. These cross-culturally consistent antonymic patterns suggest a foundational role in developing multimodal language signals.

Facial expressions are crucial for effective human and primate social communication (Bliss-Moreau & Moadab, 2017). What motivates the mapping of specific face movements on social signals, and how did they become integrated with speech? Current evidence suggests that facial expressions of *emotion* evolved from physiologically-relevant facial cues. For example, FEAR facial expressions typically involve widened eyes that increase the visual field. Conversely, DISGUST facial expressions typically involve contracted eyes and nostrils that reduce sensory input, protecting the expresser (Darwin, 1872; Susskind et al., 2008).

In line with these existing theories, we hypothesized that *social* facial expressions, such as THINKING, INTERESTED, BORED and CONFUSED and those accompanying speech, would also comprise mappings between contrasting expansion and contraction facial movements and broad social information (e.g., affective, pragmatic, semantic). To test this, we first analyzed two complementary sets of facial

expression models derived from Western European and East Asian culture participants using a highly-powered perception-based data-driven method (Yu et al., 2012): the six basic emotions (HAPPY, SURPRISE, FEAR, DISGUST, ANGER and SAD, N=30 per culture, 31 females, M_{age}=22 years; Jack et al., 2012) and four key social messages (THINKING, INTERESTED, BORED and CONFUSED, N=20 per culture, 20 females, M_{age}=22 years, Chen et al., 2020). Using non-parametric permutation testing (p<0.05), we found that, in each culture, both sets of facial expressions systematically comprise expansion and contraction facial movements (e.g., wide opened eyes, wrinkled nose, respectively) that map onto broad affective information (see Fig. 1A). Specifically, *expansion* facial movements are primarily associated with high arousal, regardless of positive/negative valence, while *contraction* facial movements are primarily associated with negative valence, regardless of high/low arousal (Nölle et al., 2021).

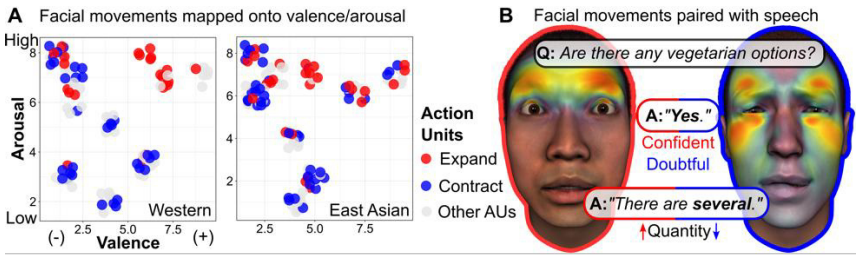


Figure 1. A: Mapping between expansion and contraction facial movements and valence and arousal, for each culture. Color-coded points denote expansion/contraction facial movements (see legend to right). In both cultures, expansion facial movements are primarily associated with high arousal, regardless of positive/negative valence (e.g., HAPPY, FEAR). Contraction facial movements are primarily associated with negative valence, regardless of high/low arousal (e.g., DISGUST, SAD, BORED, CONFUSED). **B:** Expansion and contraction facial movements also influence the interpretation of spoken words. Left: Expansion facial movements convey speaker CONFIDENCE when answering questions ("*yes*") and LARGER quantities when speakers use vague quantifiers ("*several*"). Right: Contraction facial movements convey DOUBT and SMALLER quantities.

Next, we examined whether expansion and contraction facial movements modify the perception of otherwise neutral speech. In two experiments (Nölle et al., 2022), participants in each culture (Western European English speakers, East Asian Mandarin speakers) rated speakers displaying expansion or contraction facial movements. In the first experiment, participants rated the confidence of a speaker answering *yes/no* to a question. Results showed that participants in both cultures rated speakers displaying expansion facial movements as more CONFIDENT, and those displaying *contraction* facial movements as more DOUBTFUL (Fig. 1B). In the second experiment, participants estimated the quantity the speaker referred to using a vague quantifier (e.g., *several*). Results showed that

participants in both cultures associated *expansion* facial movements with HIGHER quantities and *contraction* facial movements with LOWER quantities (Fig. 1B).

In sum, we found that social facial expressions comprise opposing expansion and contraction facial movements that map onto broad affective, semantic and pragmatic information (Bavelas & Chovil, 2018). Our results support the hypothesis that physiologically-rooted facial movements underpin a broad spectrum of facial expressions—a pattern consistent across two distinct cultures with known differences in facial expression perception (Jack, 2013). This suggests that such facial movements may have guided the development of semantic and pragmatic facial signals that support spoken language in multimodal communication (Vigliocco et al., 2014; Kendon, 2017; Holler & Levinson, 2019). Future research should examine whether similar mappings exist amongst other cultures, sign languages, and species for homologous and non-homologous expressions.

Acknowledgements

This work was supported by the National Science Centre of Poland under the agreement UMO 2023/49/B/HS2/02222, the European Research Council Grant (ERC, 759796), Wellcome Trust Grant (107802/Z/15/Z), MSCA Postdoctoral Fellowship (UKRI HORIZON-MSCA-2021-PF01), Economic and Social Research Council (ES/P000681/1), the Leverhulme Trust (Early Career Fellowship, ECF-2020-401), the University of Glasgow (Lord Kelvin/Adam Smith Fellowship, 201277), and the Chinese Scholarship Council (201306270029).

References

Bavelas, J., & Chovil, N. (2018). Some pragmatic functions of conversational facial gestures. *Gesture*, *17*(1), 98–127. https://doi.org/10.1075/gest.00012.bav

Bliss-Moreau, E., & Moadab, G. (2017). The faces monkeys make. In J. A. Russell & J. M. Fernández Dols (Eds.), *The science of facial expression* (pp. 152–171). Oxford University Press.

Chen, C., Garrod, O. G. B., Ince, R. A. A., Foster, M. E., Schyns, P. G., & Jack, R. E. (2020). Building Culturally-Valid Dynamic Facial Expressions for a Conversational Virtual Agent Using Human Perception. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 1–3. https://doi.org/10.1145/3383652.3423913

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals by Charles Darwin*. Murray.

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

Jack, R. E. (2013). Culture and facial expressions of emotion. *Visual Cognition*, *21*(9–10), 1248–1286. https://doi.org/10.1080/13506285.2013.835367

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*(19), 7241–7244. https://doi.org/10.1073/pnas.1200155109

Kendon, A. (2017). Reflections on the "Gesture-First" hypothesis of language origins. *Psychonomic Bulletin & Review*, *24*(1), 163–170. https://doi.org/10.3758/s13423-016-1117-3

Nölle, J., Chen, C., Hensel, L. B., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2021). Facial expressions of emotion include iconic signals of rejection and acceptance. *Journal of Vision*, *21*(9), 2932. https://doi.org/10.1167/jov.21.9.2932

Nölle, J., Wu, Y., Arias, P., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2022). More Than Affect: Human Facial Expressions Provide Iconic and Pragmatic Functions. In A. Ravignani, R. Asano, D. Valente, F. Ferretti, S. Hartmann, M. Hayashi, Y. Jadoul, M. Martins, Y. Oseki, E. D. Rodrigues, O. Vasileva, & S. Wacewicz (Eds.), *Proceedings of the Joint Conference on Language Evolution (JCoLE)*. https://doi.org/10.17617/2.3398549

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

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651). https://doi.org/10.1098/rstb.2013.0292

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

The aesthetic appeal of sign languages: intrinsic appeal or cultural associations?

Theresa Matzinger[*1,2], Monika Boruta[2,3], Marek Placiński[3], Aoifе Jordan[1] and Hannah Marwan-Schlosser[1]

[*]Corresponding Author: theresa.matzinger@univie.ac.at
[1]Department of English and American Studies, University of Vienna, Vienna, Austria
[2]Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria
[3]Centre for Language Evolution Studies, Nicolaus Copernicus University Toruń, Toruń, Poland

We explore how socio-cultural associations influence as how appealing American (ASL) and French (LSF) Sign Language are perceived. We hypothesize that LSF will be rated as more appealing than ASL, but only when participants know which languages they are rating, reflecting a "Latin-lover" effect seen in spoken languages. Additionally, we investigate how the meaning of signs affects their perceived appeal. Findings contribute to understanding how aesthetic appeal and cultural associations may shape language evolution in both signed and spoken systems.

Are some languages perceived as more appealing than others, and if so, why? While some research suggests intrinsic differences in languages' appeal (e.g., Kogan & Reiterer 2021; Winkler et al. 2023), others argue that appeal is largely socio-culturally constructed (Anikin et al. 2023). Prior research has explored those questions for spoken languages, but similar endeavors are missing for sign languages. We address this gap by investigating if sign languages, specifically American (ASL) and French (LSF) Sign Language, are perceived differently in terms of appeal and how socio-cultural associations shape these perceptions.

Previous studies on spoken languages have found a "Latin-lover" effect, where Romance languages are perceived as more beautiful than, e.g., Germanic or Slavic languages (Kogan & Reiterer, 2021; Reiterer et al., 2020; Burchette, 2014). This effect likely stems from the fact that many (largely WEIRD) study participants were familiar with Romance languages, possibly triggering positive cultural associations. Based on this, we hypothesize that when participants are unaware of the languages being rated, ASL and LSF will be perceived as equally appealing. However, when participants are informed about the languages, we predict that LSF will be rated as more appealing than ASL, reflecting cultural

associations that may drive the "Latin-lover" effect for spoken languages.

As a secondary aim, drawing on the iconic potential of sign languages (e.g. Mineiro et al., 2017; Pizutto & Volterra, 2000), we tested if the meaning of a sign influences its perceived appeal and valence (cf. Louverse & Qu 2017). In our experiment, half of the signs convey positive (e.g., love, peace) and the other half negative concepts (e.g., hate, war). We expect that signs representing positive concepts will be rated as more appealing than those representing negative ones, even though participants will be unaware of their meanings.

In our experiment, 200 WEIRD participants who are unfamiliar with sign languages or deaf culture are presented with short, standardized video clips of an experimenter performing signs for the same 20 concepts (10 being positively and 10 being negatively valenced) in both ASL and LSF. The signer's face is blurred to remove facial cues. Participants rate each sign on two Likert scales: (1) how appealing they find the look of the signs and (2) how positive they perceive the meaning of the signs. One half of the participants is uninformed, while the other half is informed whether the presented signs are part of either ASL or LSF.

Preliminary results of a cumulative link mixed model for the uninformed group (n=40) show, as predicted, no significant differences in the intrinsic appeal of ASL and LSF. However, signs representing positive concepts received significantly higher ratings for both appeal and valence compared to those representing negative concepts (Fig. 1). This suggests that even without knowledge of the language, participants can differentiate signs based on their emotional content, and that perceived valence correlates with perceived appeal.

We will complement these initial findings with data from the informed participant group, predicting that this group will perceive the two languages differently in terms of appeal. This research offers new insights into how sign languages are perceived aesthetically and the extent to which socio-cultural associations influence these perceptions. The perception of linguistic aesthetics can shed light on the cognitive and social processes involved in language formation and change. Additionally, the influence of cultural associations on linguistic perceptions may reflect broader patterns of language evolution, where certain features are more likely to be favored or maintained due to their emotional or social appeal (Matzinger et al. 2021). By examining these processes in the context of sign languages, we can gain a deeper understanding of the evolutionary forces shaping both spoken and signed communication systems.

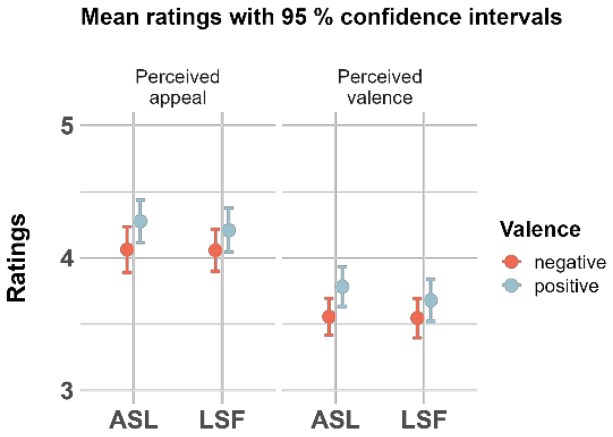


Figure 1. Participants' appeal and valence ratings of negative and positive concepts in ASL and LSF (rating scale from 1-7). Participants were not informed about the languages or meanings of the signs.

Acknowledgements

Theresa Matzinger was supported by a Disruptive Innovation Grant from the Austrian Academy of Sciences and the Austrian Science Fund (Grant number: DI_2023-108_MATZINGER_BEALP).

References

Anikin, A., Aseyev, N., & Erben Johansson, N. (2023). Do some languages sound more beautiful than others? *Proceedings of the National Academy of Sciences*, *120*(17), e2218367120. https://doi.org/10.1073/pnas.2218367120 Burchette, J. (2014). Sexiest accents poll: Where do people have the voice of seduction? CNN International. Retrieved from https://edition.cnn.com/travel/article/sexy-accents/index.html (30.01.2025)

Kogan, V. V., & Reiterer, S. M. (2021). Eros, Beauty, and Phon-Aesthetic Judgements of Language Sound. We Like It Flat and Fast, but Not Melodious. Comparing Phonetic and Acoustic Features of 16 European Languages. *Frontiers in Human Neuroscience*, *15*, 578594. https://doi.org/10.3389/fnhum.2021.578594

Louwerse, M., & Qu, Z. (2017). Estimating valence from the sound of a word: Computational, experimental, and cross-linguistic evidence. Psychonomic Bulletin & Review, 24, 849-855. https://link.springer.com/article/10.3758/s13423-016-1142-2

Matzinger, T., Specker, E., Ritt, N., & Fitch, W. T. S. (2021). Aesthetic perception of prosodic patterns as a factor in speech segmentation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*, 2992-2998. https://escholarship.org/uc/item/4tm94762

Mineiro, A., Carmo, P., Caroça, C., Moita, M., Carvalho, S., Paço, J., & Zaky, A. (2017). Emerging linguistic features of Sao Tome and Principe Sign Language. *Sign Language & Linguistics, 20/1,* 109-128. https://doi.org/10.1075/sll.20.1.04min

Pizzuto, E., & Volterra, V. (2000). Iconicity and transparency in sign languages: A cross-linguistic cross-cultural view. In K. Emmorey & H. Lane (Eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (pp. 261–286). Lawrence Erlbaum Associates Publishers. https://psycnet.apa.org/record/2000-03924-010

Reiterer, S. M., Kogan, V., Seither-Preisler, A., & Pesek, G. (2020). Foreign language learning motivation: Phonetic chill or Latin lover effect? Does sound structure or social stereotyping drive FLL? In K. D. Federmeier & H.-W. Huang (Eds.), *The psychology of learning and motivation: Adult and second language learning* (pp. 165–205). Elsevier Academic Press. https://doi.org/10.1016/bs.plm.2020.02.003

Winkler, A., Kogan, V. V., & Reiterer, S. M. (2023). Phonaesthetics and personality—Why we do not only prefer Romance languages. *Frontiers in Language Sciences*, *2*. https://www.frontiersin.org/articles/10.3389/flang.2023.1043619

The domestication of speech

Antonio Benítez-Burraco*[1] , Raffaela Lesch[2]

*Corresponding Author: abenitez8@us.es

[1]Department of Spanish, Linguistics, and Theory of Literature, Faculty of Philology, University of Seville, Seville, Spain

[2] Department of Biology, Donaghey College of Science, Technology, Engineering, and Mathematics, University of Arkansas, Little Rock, USA

This paper connects two theories of language evolution that have developed in parallel: one linking the emergence of modern speech to changes in the speech organs, and the other linking language complexification to changes in human behavior and culture. Our contention is that selected changes in the speech organs accounting for modern speech, particularly, in the hyoid bone, might have resulted (or being favored) by our evolutionary trend towards a more prosocial phenotype (aka human self-domestication), which complexified language through a cultural mechanism. If correct, this approach would offer a more parsimonious explanation of modern speech evolution and make hypotheses about speech-like abilities in other hominin species more testable.

Introduction

Language evolution studies have flourished during the last years. Language does not fossilize, but its evolution can be inferred from indirect evidence, sometimes known as "proxies", "windows" or "fossils" of language (e.g. Botha, 2016). In this paper, we focus on the evolution of modern speech, a favorite topic in the language evolution literature. Speech is the usual way in which humans exteriorize their linguistic thoughts, it is easy to analyze (in contrast to e.g. language processing by the brain), and there are true fossil remains of some of the speech organs, like parts of the hard palate, or the hypoglossal canal. As with most proxies or windows, the consideration of these remains has resulted in different, sometimes opposite views of the emergence of speech (e.g. the reconstructions of the Neanderthal vowel space by Lieberman and Crelin, 1971 vs Boë and colleagues, 2002). At the same time, after decades of extensive research on language evolution, most researchers would now agree that language did not emerge from scratch because of some abrupt change in our biology, but from a myriad of changes that gradually impacted on our body, behavior, and culture. A corollary is that any fruitful hypothesis about language evolution must be

grounded on a solid narrative of the evolution of the human species. Our aim in the paper is to provide an evolutionary rationale for the attested changes in our speech organs that resulted in modern speech, and ultimately, to contribute to tie together two narratives of language evolution that have run in parallel for the most time: the evolution of speech via biological changes, and the sophistication of languages via a cultural mechanism. More specifically, we will examine the potential effect on the hyoid bone, which is a robust anatomical proxy for speech evolution, of our purported self-domestication, which reshaped our behavior and culture mostly.

The hyoid bone and the evolution of speech

The hyoid bone is a floating bone that provides anchoring to muscles of the floor of the mouth and the tongue, and the larynx, contributing to the fine-tuning of tongue movements, and ultimately, to the diversity of speech sounds. The hyoid bone has changed notably in the genus *Homo*. The Neanderthal hyoid was like ours (Arensburg et al., 1989), with similar internal architectures and micro-biomechanical capacities (D'Anastasio et al., 2013). By contrast, the hyoid from *Homo erectus* shows some archaic features and differences with the human and the Neanderthal hyoids, which could be indicative of a reduced capability for modulating the length of the vocal tract, and accordingly, of differences in their speech-like abilities (Capasso et al., 2008).

The human self-domestication (HSD) view of language evolution

The HSD hypothesis argues that the human phenotype is the outcome of an evolutionary process similar to animal domestication. Domestication involves an initial selection for tameness and tolerance to contact with humans, but interestingly, it usually brings about a set of co-occurring distinctive traits—physical, cognitive, and behavioral: the domestication syndrome (Wilkins et al., 2014; see Sanchez-Villagra et al., 2016 and Lord et al., 2020 for critical views). Most of these features are observed in modern humans compared to extant primates and extinct hominins (Shea, 1989; Leach, 2003; Somel et al., 2009; Zollikofer and Ponce de León, 2010; Plavcan, 2012; Fukase et al., 2015; Stringer, 2016). The behavioral changes (and to some extent the cognitive and physical modifications) brought about by HSD have been hypothesized to have favored the evolution of human-distinctive traits, such as our enhanced social cognition and increased cooperation, and ultimately, our sophisticated culture and advanced technology (Hare, 2017; Hare and Woods, 2020). Benítez-Burraco and Progovac (2020) have further suggested that also language evolved under the effects of HSD, gaining structural complexity and functional versatility with time. This sophistication was seemingly achieved through a cultural process that was favoured by our increased HSD, including increased contacts between people and extended playing behavior, which are mechanisms that promote the

complexification of languages through a cultural process (Benítez-Burraco and Kempe, 2018; Langley et al. 2020).

Linking changes in the hyoid bone to HSD

Wilkins and colleagues (e.g. 2014, 2021) have hypothesized that the co-occurrence of similar traits in most domesticates can be explained if attenuated aggression responses (and ultimately, low stress levels), as found in domesticated animals, reduce the input to the neural crest, an embryonic structure that supports the ontogenetic development of numerous body parts (see again; Sánchez-Villagra et al., 2016 and Lord et al., 2020 for critical views). HSD might have entailed similar changes in neural crest function, even if it was triggered by external factors (see Benítez-Burraco, 2025 for a recent review), since neurocristopathies in humans result in clinical features that resemble traits found in domesticates (Sánchez-Villagra et al., 2016). Interestingly, the voice-producing structures, specifically the larynx and hyoid, depend on neural crest cells for correct formation during embryonic development (Tabler et al. 2017). A recent publication by Lesch and Fitch (2024) suggests that in domesticates, fewer neural crest cells arrive at target sites relevant to the formation of these voice-producing structures, this resulting in an overall smaller larynx and hyoid. This would mean that, corrected for body size, and despite species specific adaptations, domesticated mammals will have a smaller larynx and hyoid bones compared to their ancestral wild population. Therefore, if humans indeed underwent HSD, our hyoid pattern changes should share certain overall patterns with other domesticated mammals across the phylogenetic tree (Fig. 1). In our contribution, we will further discuss an analysis pipeline aimed at comprehensively testing the putative role of the hyoid a proxy of HSD via the Neural Crest Domestication Syndrome hypothesis. Linking the changes in speech organs/proxies like the hyoid bone to HSD offers a more parsimonious explanation of modern speech evolution and makes hypotheses about speech in other hominin species more testable.

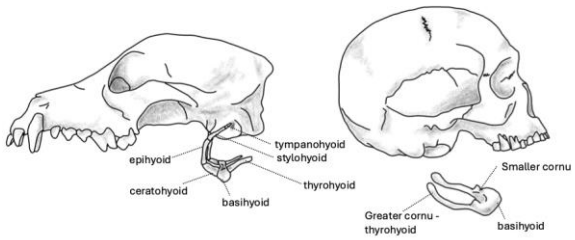


Figure 1. Illustration of the hyoid bone and apparatus in a dog (*Canis familiaris*) and a human (*Homo sapiens*).

References

Arensburg, B., Tillier, A. M., Vandermeersch, B., Duday, H., Schepartz, L. A., & Rak, Y. (1989). A Middle Palaeolithic human hyoid bone. *Nature, 338(6218),* 758–760. https://doi.org/10.1038/338758a0

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. (2025). How (and why) languages became more complex as we evolved more prosocial: the human self-domestication view. *Frontiers in psychology, 15,* 1499994. https://doi.org/10.3389/fpsyg.2024.1499994

Benítez-Burraco, A., & Kempe, V. (2018). The Emergence of Modern Languages: Has Human Self-Domestication Optimized Language Transmission?. *Frontiers in psychology, 9,* 551. https://doi.org/10.3389/fpsyg.2018.00551

Boë, L.-J., Heim, J.-L., Honda, K., & Maeda, S. (2002). The potential Neanderthal vowel space was as large as that of modern humans. *Journal of Phonetics, 30,* 465–484. doi:10.1006/jpho.2002.0170.

Botha, R. P. (2016). *Language evolution: The windows approach.* Cambridge University 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.

D'Anastasio, R., Wroe, S., Tuniz, C., Mancini, L., Cesana, D. T., Dreossi, D., Ravichandiran, M., Attard, M., Parr, W. C., Agur, A., & Capasso, L. (2013). Micro-biomechanics of the Kebara 2 hyoid and its implications for speech in Neanderthals. *PloS one, 8(12),* e82261. https://doi.org/10.1371/journal.pone.0082261

Fukase, H., Kondo, O., & Ishida, H. (2015). Size and placement of developing anterior teeth in immature Neanderthal mandibles from Dederiyeh Cave, Syria: implications for emergence of the modern human chin. *American Journal of Physical Anthropology 156,* 482-488. Doi: 10.1002/ajpa.22665.

Hare, B. (2017). Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality. *Annual review of psychology, 68,* 155–186. https://doi.org/10.1146/annurev-psych-010416-044201

Hare, B., Woods, V. (2020). *Survival of the friendliest: Understanding our origins and rediscovering our common humanity*. Random House, NY.

Langley, M. C., Benítez-Burraco, A., & Kempe, V. (2020). Playing with language, creating complexity: Has play contributed to the evolution of complex language?. *Evolutionary anthropology, 29(1),* 29–40. https://doi.org/10.1002/evan.21810

Leach, H. (2003). Human domestication reconsidered. *Current anthropology, 44(3),* 349-368.

Lesch, R., & Fitch, W. T. (2024). The domestication of the larynx: The neural crest connection. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution, 342(4),* 342–349. https://doi.org/10.1002/jez.b.23251

Lieberman, P., & Crelin, E. S. (1971). On the speech of Neanderthal man. *Linguistic Inquiry, 2*, 203– 222.

Lord, K. A., Larson, G., Coppinger, R. P., & Karlsson, E. K. (2020). The History of Farm Foxes Undermines the Animal Domestication Syndrome. *Trends in ecology & evolution, 35(2),* 125–136. https://doi.org/10.1016/j.tree.2019.10.011

Plavcan, J. M. (2012). Sexual size dimorphism, canine dimorphism, and male-male competition in primates. *Human Nature, 23,* 45–67.

Sánchez-Villagra, M. R., Geiger, M., & Schneider, R. A. (2016). The taming of the neural crest: a developmental perspective on the origins of morphological covariation in domesticated mammals. *Royal Society open science, 3(6),* 160107. https://doi.org/10.1098/rsos.160107

Shea, B (1989). Heterochrony in human evolution: the case for neoteny reconsidered. *American Journal of Physical Anthropology, 32,* 69–101.

Somel, M., Franz, H., Yan, Z., Lorenc, A., Guo, S., Giger, T., Kelso, J., Nickel, B., Dannemann, M., Bahn, S., Webster, M. J., Weickert, C. S., Lachmann, M., Pääbo, S., & Khaitovich, P. (2009). Transcriptional neoteny in the human brain. *Proceedings of the National Academy of Sciences of the United States of America, 106(14),* 5743–5748. https://doi.org/10.1073/pnas.0900544106

Stringer, C. (2016). The origin and evolution of Homo sapiens. *Philosophical Transactions of the Royal Society B Biological Sciences, 371,* 20150237.

Tabler, J. M., Rigney, M. M., Berman, G. J., Gopalakrishnan, S., Heude, E., Al-Lami, H. A., Yannakoudakis, B. Z., Fitch, R. D., Carter, C., Vokes, S., Liu, K. J., Tajbakhsh, S., Egnor, S. R., & Wallingford, J. B. (2017). Cilia-mediated Hedgehog signaling controls form and function in the mammalian larynx. *eLife, 6,* e19153. https://doi.org/10.7554/eLife.19153

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. https://doi.org/10.1534/genetics.114.165423

Wilkins, A. S., Wrangham, R., & Fitch, W. T. (2021). The neural crest/domestication syndrome hypothesis, explained: reply to Johnsson, Henriksen, and Wright. *Genetics, 219(1),* iyab098. https://doi.org/10.1093/genetics/iyab098

Zollikofer, C. P., & Ponce de León, M. S. (2010). The evolution of hominin ontogenies. *Seminars in cell & developmental biology, 21(4),* 441–452. https://doi.org/10.1016/j.semcdb.2009.10.012

The effect of posture on the production of multimodal communication

Susanne Fuchs[*1], Mingtong Li[2], Jiahao Yang[2,3], and Suzanne Aussems[2]

[*]Corresponding Author: fuchs@leibniz-zas.de
[1]Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany
[2]Department of Psychology, University of Warwick, Coventry, UK
[3]Institute for Digital Security and Behaviour (IDSB), University of Bath, Bath, UK

Abstract

Humans maintain upright posture on a minimal base of support, relying on continuous postural control to stay balanced. We propose that the evolution of higher-level postural control, driven by bipedalism, shaped multimodal communication. To test this idea, we experimentally manipulated adults' postural stability to assess how it affects speech and gesture production in an interactive Taboo game. Participants communicated with a partner while standing on solid ground or on a wobble board, increasing demands on their postural control during communication.

1. Introduction

Maintaining an upright posture requires constant neuromuscular control to stabilize the body against gravity on a minimal base of support. We propose that these postural mechanisms not only support balance but also contribute to multimodal communication production.

Evidence across multiple domains demonstrates that articulators involved in posture overlap with those used in speech and gesture. For example, clinical studies reveal mutual influences between jaw-head motion and body posture (Alghadir, Zafar, & Iqbal, 2015; Sakaguchi et al., 2007; Yamabe, Yamashita, & Fujii, 1999). Furthermore, kinetic studies show that upper limb gestures alter the center of pressure and postural stability (Pouw, Raphael, Burchardt, & Selen, 2025) and may affect acoustic speech signals (Pouw, Paxton, Harrison, & Dixon, 2020). Furthermore, a recent machine learning study reveals that vertical gesture kinetics predict speech prosody (Momsen & Coulson, 2025). Together, these findings suggest that speaking, gesturing, and postural control are biomechanically interconnected processes.

Human communication involves overlapping muscle systems that are more interconnected than traditionally recognized. While respiratory and laryngeal muscles handle speech production (Simonyan & Horwitz, 2011), gestural muscles include secondary respiratory muscles–abdominals, back muscles, and shoulder girdle–that also support vocalization (Pouw et al.,

2025). We propose that evolutionary development of sophisticated postural control, following the transition from quadrupedalism to bipedalism, was crucial for refining coordination between these systems in multimodal communication. Although bipedalism predates language, early hominins possessed only rudimentary postural control (Ruff, 2015). The precise postural stability necessary for coordinated speech–gesture communication developed gradually throughout human evolution. Thus, while bipedalism did not directly give rise to language, it served as a precursor that prompted anatomical and neurological adaptations, laying the groundwork for the coordinated vocal and gestural systems that underpin human multimodal communication.

2. Methodology

To test this idea, we manipulated adults' postural stability by examining how standing on stable versus unstable surfaces influenced multimodal communication during an interactive Taboo game. Players attempted to get their partner to guess target words without using specific "taboo" words (e.g., for "apple," taboo words included "fruit," "red," "tree," "pie," and "eat"). While players were restricted from saying target or taboo words, they could use other verbal descriptions and were encouraged to use gestures, facial expressions, and body language.

Sixty pairs of adult participants ($N = 120$) played the game while standing on solid ground (stable condition) or on a wobble board (unstable condition). Within each dyad, participants took turns giving clues and guessing, and clue-givers alternated between standing on the solid ground and wobble, following a within-subjects design.

3. Predictions

We formulated several predictions for clue-givers' speech and gesture production as a function of postural stability.

For speech production, we expect slower speech, increased pausing, and reduced articulation on the wobble board compared to solid ground. Acoustic analysis using PRAAT (Boersma & Weenink, 2002) will allow us to measure speech rate, pausing, fundamental frequency (f0), amplitude, and peaks of the amplitude envelope.

For gesture production, we expect reduced gesture frequency due to increased postural demands. When gestures occur, we hypothesize they will be produced closer to the center of gravity, favor bilateral over unilateral movements, and exhibit greater symmetry to minimize balance disruption. Movement analysis using pose2sim (Pagnon, Domalain, & Reveret, 2021) will allow us to quantify the kinematics of the center of mass, velocity, and upper extremity.

This study will provide original empirical evidence for how postural control shapes multimodal communication, contributing to theories of language evolution.

Acknowledgements

This work was supported by a small research grant from The British Academy/Leverhulme Trust awarded to Suzanne Aussems (reference number: SRG2324\240765).

References

Alghadir, A. H., Zafar, H., & Iqbal, Z. A. (2015). Effect of three different jaw positions on postural stability during standing. *Functional Neurology*, *30*(1), 53.

Boersma, P., & Weenink, D. (2002). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Momsen, J. P., & Coulson, S. (2025). Decoding prosodic information from motion capture data: The gravity of co-speech gestures. *Open Mind: Discoveries in Cognitive Science*, *9*, 652–664.

Pagnon, D., Domalain, M., & Reveret, L. (2021). Pose2sim: An end-to-end workflow for 3d markerless sports kinematics—part 1: Robustness. *Sensors*, *11*(19), 6530–6557. (Version 0.3.6 released 2023-05-05)

Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proceedings of the National Academy of Sciences*, *117*(21), 11364–11367.

Pouw, W., Raphael, W., Burchardt, L. S., & Selen, L. P. (2025). The human voice aligns with whole-body kinetics. *Proceedings of the Royal Society: Biological Sciences*, *292*(2047), 20250160.

Ruff, C. B. (2015). Body, evolution of. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences (second edition)* (Second Edition ed., p. 723-727). Oxford: Elsevier.

Sakaguchi, K., Mehta, N. R., Abdallah, E. F., Forgione, A. G., Hirayama, H., Kawasaki, T., & Yokoyama, A. (2007). Examination of the relationship between mandibular position and body posture. *CRANIO®*, *25*(4), 237–249.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *17*(2), 197–208.

Yamabe, Y., Yamashita, R., & Fujii, H. (1999). Head, neck and trunk movements accompanying jaw tapping. *Journal of Oral Rehabilitation*, *26*(11), 900–905.

The Emergence of Lexical Items through Demonstration: Evidence from Iconic Expressions

Koji Kawahara*[1]

*Corresponding Author: kkoji@nufs.ac.jp
[1]Department of British and American Studies, Nagoya University of Foreign Studies, Japan

This paper discusses how lexical items emerge through demonstration, focusing on iconic linguistic signs, particularly interjections and ideophones. These items illustrate how expressive, non-arbitrary signs enter the lexicon through demonstrative usage. Demonstration allows interjections, ideophones and loanwords to become grammatically integrated, forming predicates and acquiring stable category labels required for syntactic merge. Thus, iconic signs initially introduced demonstratively can evolve into fully established lexical items, highlighting the role of demonstration in lexical emergence and grammatical integration.

1. Introduction

Human language is typically characterised by an extensive lexicon coupled with a relatively limited set of functional categories. Unlike pidgin-based creole languages, which inherit their extensive vocabulary from source languages (e.g. the lexifier (Mufwene, 2001))(Bickerton, 1981; Aboh, 2015), emerging languages such as Nicaraguan Sign Language provide a compelling example of how lexicons develop without direct linguistic predecessors. One prominent feature observed in these cases is iconicity; signs often arise through imitation, capturing meaning by mimicking their referents. Motamedi et al. (2019) define iconicity as the perceived resemblance or motivated relationship between the form of a linguistic signal and its meaning. It is not limited to ideophones but is seen as a widespread feature across modalities (spoken, signed and gestural languages). According to them, there are two key dimensions of iconicity: operational iconicity and functional iconicity. The former is about perceptual similarity, the extent to which the form resembles meaning. The latter concerns how usable the signal is for guessability or learning, the extent to which a naive listener infers the meaning from the form. The following table is a summary of their discussion.

Aspect	Summary
Definition	Resemblance or motivated link between form and meaning
Two Types	*Operational*: perceptual similarity *Functional*: ease of guessing meaning
Measurement Methods	1. Intuition-based (e.g., expert coding) 2. Behavioural (e.g., rating, guessing, production) 3. Data-driven (e.g., corpus/statistical patterns)
Key Insight	Iconicity is multi-dimensional and context-dependent
Purpose of Toolbox	To help researchers choose appropriate methods

Winter et al. (2024) argue that iconicity is a widespread and variable feature of the lexicon. Their point is summarised in the following table.

Feature	Iconicity	Systematicity	Arbitrariness
Definition	Form resembles meaning	Form predicts class/category	No link between form and meaning
Mapping Type	Motivated, intuitive	Probabilistic, statistical	Conventional, random
Measurement	Rating tasks (e.g., 1-7 scale)	Corpus/statistical models	Residual category
Examples	*buzz*, *wiggle*	*gl-* (e.g., *glow*)	*dog*, *chair*
Language Modality	Strong in sign/sound-symbolic languages	Common in morphology/syntax	Typical in general vocabulary

This talk discusses the nature of grammatical integration driven by iconicity, arguing that lexical items forming natural classes become targets for grammatical operations due to their categorical identities.

2. Symbolic items

It is generally assumed that most lexical items are arbitrary and conventional, despite acknowledging the existence of sound symbolism. According to Knoeferle et al. (2017), sound symbolism reflects a motivated relationship between sound and meaning, structured by specific acoustic cues that differ depending on the semantic dimension involved (e.g., size vs. shape). However, lexical items with iconicity present interesting cases for linguistic analysis.

2.1. *Different modules*

Following Davidson (2015), demonstration can provide insights into the interpretation of symbolic items. Demonstrative expressions, introduced by constructions such as *be like*, illustrate how non-verbal and expressive aspects of language contribute to semantic interpretation, as shown in the examples below:

(1) a. My cat was like “feed me!”
b. Bob saw the spider and was like “ahh! [in a scared voice].”

An advantage of demonstration theory is that it explicitly captures expressive nuances. In (1a), the cat's imagined utterance is conveyed in human language, interpreting the cat's behavior through linguistic means. Similarly, the exclamation *ahh!* in (1b) communicates Bob's emotional state through vocal tone and expression. Demonstrative constructions can also incorporate non-verbal signals such as gestures and facial expressions, indicating their role in the emergence of lexical meaning. These demonstrative expressions reflect cross-modality integration in lexical development and offer valuable perspectives on language evolution.

A notable class of symbolic linguistic items are ideophones, which are particularly abundant in languages across Asia, Africa, and the Americas (Akita and Dingemanse, 2019). In Japanese, ideophones commonly function as adverbs, often optionally marked by the quotative particle *-to*, which typically does not affect their semantic interpretation. However, *-to* becomes obligatory when ideophones are employed in non-idiomatic contexts. Furthermore, ideophones or phonomimes like *zaazaa* ('pouring heavily') show stronger associations with auditory contexts (e.g., *kikoeru* 'audible'), but have a more indirect relationship with visual perception, as illustrated below (Akita and Usuki, 2016):

(2) a. Ame-ga *zaazaa(-to)* hut-te iru no-ga kikoeru.
rain-NOM IDEO-QUOT fall-CONJ be NMLZ-NOM be.audible
'I hear the rain pouring heavily.'

b. Ame-ga *zaazaa(??-to)* hut-te iru no-ga mieru.

This indicates that the quotative particle *-to* interacts closely with modality-specific sensory interpretations, reflecting the iconic origins of such expressions (Kawahara, 2022). Demonstrative expressions, by themselves, have some illocutionary force. Like the exclamation *ahh!*, the interjections such as *yes, no, huh, eh* can stand alone to have some force. Hence, they are the primitives of auditory signs and cannot be categorised, or grammatically inflected.

2.2. *Ideophones*

The fundamental usage of ideophones is typically holophrastic or adverbial; however, a considerable number of ideophones also function as predicates. Predicative ideophones primarily belong to the category of phenomimes, describing states, motion, or psychological conditions, as illustrated below:

(3) a. *wanwan-suru '(lit.) bowwow-do', #gangan-suru "(intended) *bang, (possible), a splitting headache"

b. subesube-suru 'smooth', burabura-suru 'stroll'

Since oral languages are fundamentally based on sound, audio-related ideophones (phonomimes) can be hypothesised to be most iconic (Dingemanse, 2012). In this context, Akita (2009) proposes that strongly iconic ideophones generally resist functioning as predicates. This claim appears justified given that phonomimes typically exhibit limited grammatical integration. I suggest the following implicational hierarchy.

(4) An inverse relation between (operational) iconicity and lexical integration: More iconic items tend to be less lexically-integrated and vice versa. (SOUND < VISUAL PATTERNS and MOVEMENT < OTHER SENSORY PERCEPTIONS and COGNITIVE STATE)

	Incorporability	**Arbitrariness**
Auditory items	Weak (NO SOUND-EMITTING VERBS)	(Strong) Link between form and meaning
Visual items	Relatively weak (e.g. kirakira-suru 'sparkle', NO MOTION VERBS)	Some link between form and meaning
Other items	Strong (e.g. dokidoki-suru 'thrilled', wakuwaku-suru 'excited')	No link between form and meaning
Demonstrative expressions	Weak	Strong link between form and meaning

2.3. *Neologism*

Japanese has a productive mechanism for verb formation through the use of the light verb *(su)-ru*. For instance, during a recent trend involving tapioca drinks, many young people coined the verb *tapi-ru* ('to enjoy tapioca'). Similar verb formations can be seen historically, such as *ryou-ru* ('to cook'), derived from the noun *ryouri* ('cooking'), which originated around 350 years ago and remains in use in certain regions of Japan. Loanwords like *agitation* and *sabotage* have also undergone this process, resulting in the verbs *agi-ru* ('to agitate') and *sabo-ru* ('to skip one's duties'), respectively. Such lexical conversions occur when loanwords are sufficiently integrated into the linguistic system. Another example includes *kisi-mu* ('to creak'), formed from the phonomime *kisikisi* ('squeaking sound').

This process demonstrates that non-iconic ideophones or frequently used loanwords can be integrated into predicate structures, thus becoming central components of sentences. Such lexical items initially gain linguistic acceptance through demonstrative usage, which allows them to enter the general lexicon and form natural categories. Through repeated use, they become categorised clearly enough to serve as targets for grammatical operations (Chomsky, 2013), solidifying their status as established lexical items. Linguistic items that are sorted out based on their common characteristics can be a target of another grammatical operation, because they are integrated into some lexical categories.

References

Aboh, E. O. (2015). *The Emergence of Hybrid Grammars*. Cambridge University Press, Cambridge, Great Britain.

Akita, K. (2009). *A Grammar of Sound-Symbolic Words in Japanese*. PhD thesis, Kobe University.

Akita, K. and Dingemanse, M. (2019). Ideophones (mimetics, expressives). In Aronoff, M., editor, *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, Oxford.

Akita, K. and Usuki, T. (2016). A constructional account of the 'optional' quotative marking on Japanese mimetics. *Journal of Linguistics*, 52:245–275.

Bickerton, D. (1981). *Roots of Language*. Language Science Press, Berlin.

Chomsky, N. (2013). Problems of projection. *Lingua*, 18:1–35.

Davidson, K. (2015). Quotation, demonstration, and iconicity. *Linguistics and Philosophy*, 38:477–520.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6:654–672.

Kawahara, K. (2022). Introducing iconicity: The semantics of ideophones and the quotative particle. In *CLS*, volume 56, pages 219–232. Chicago Linguistics Society.

Knoeferle, K., Li, F., Spence, C., and Winter, B. (2017). What drives sound symbolism? different acoustic cues underlie sound–size and sound–shape mappings. *Scientific Reports*, 7(1):1–11.

Motamedi, Y., Perlman, M., Ortega, G., and Vigliocco, G. (2019). The iconicity toolbox: Empirical approaches to measuring iconicity. *Language and Cognition*, 11(2):188–212.

Mufwene, S. S. (2001). *The Ecology of Language Evolution*. Cambridge University Press, Cambridge, Great Britain.

Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., and Perlman, M. (2024). Iconicity ratings for 14,000+ English words. *Behavior Research Methods*, 56(3):1640–1655.

The evolution of imagination: A neural transition from reptiles to mammals

Oryan Zacks

oryan.zacks@gmail.com

The Cohn Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, Tel Aviv, Israel.

Imagination, the ability to generate mental simulations, is a key cognitive capacity underlying planning, problem-solving, and, critically, language. Despite its centrality in human cognition, the evolutionary origins of imagination remain underexplored. This paper argues that imagination emerged in early mammals, enabling complex behaviors such as episodic-like memory, inference, and theory of mind. This evolutionary transition was facilitated by a significant development in the neocortex and hippocampus, which together form the core neural network supporting imagination in all mammals, including humans.

Human imagination is a complex, rich and highly developed cognitive ability, constituting an important part of human experience. It is a critical component of human linguistic abilities and was probably an evolutionary pre-requisite to developing symbolic language (Corballis, 2019). Furthermore, a well-functioning imagination is critical for mental health and well-being, and foundational for creativity and empathy. Nevertheless, imagination is critically understudied as such, with little systematic, in-depth, or interdisciplinary research addressing it directly. Adopting an evolutionary approach, this paper presents a comparative study of imagination across mammals and reptiles, considering both behavioral evidence and differences between neural organization in these two lineages.

According to Tomasello (2014, p. 9), "imagining is nothing more or less than the "off-line" simulation of potential perceptual experiences." This definition can theoretically apply to non-human animals, although the major challenge lies in measuring such mental simulations in the absence of verbal report. Deliberate planning is suggested here as a strong, testable behavioral marker of off-line simulations in non-human animals. Additional behaviors that may support the claim that non-human animals can and do imagine include theory of mind, inferential and causal reasoning, episodic-like memory, and dreaming. These behaviors are most robustly documented in great apes, with sparser evidence in other mammalian species such as monkeys, rodents, and bats. In contrast, no

published results demonstrate these behaviors in reptiles, hinting at a qualitative difference in cognitive abilities between reptiles and mammals.

A core neural network supporting imagination has been identified in the human brain (Schacter et al., 2012), which overlaps significantly with the Default Mode Network (DMN; Raichle, 2015; Schacter & Addis, 2020). Interestingly, this network spans two different cortical architectures: the hippocampus and association areas belonging to the neocortex. Understanding the evolution of these two neural architectures is critical in understanding their function and involvement in complex cognition, including imagination and language. While both neocortex and hippocampus have homologs in reptiles, they underwent significant expansion and reorganization in the base of the mammalian lineage.

The reptilian medio-dorsal pallium consists of a thin layer of interconnected principal excitatory cells. In mammals, both the neocortex and hippocampus increased in size, neuron numbers and neuron density. This was achieved by transforming the original reptilian 2D sheet of neurons into a 3D structure. In the neocortex, this meant stacking multiple layers of neurons, while the hippocampus folded upon itself to create its characteristic S-shape. These two structures also gained new organizational principles that form the basis of the canonical cortical column microcircuit and the hippocampal trisynaptic loop. This architectural shift was accompanied by the emergence of new cell types, notably large pyramidal neurons with multiple, distinct dendritic zones. The resulting modularity of the neocortical architecture opened an opportunity for further evolutionary changes, seen in the multiple mammalian lineages in which neocortical territories proliferated and gained new functions.

Behavioural and neural findings point to a significant cognitive transition between reptiles and mammals. While it is recognized that mammals vary widely in their imaginative abilities, they all seem to have a form of *minimal* imagination, absent in reptiles. This ability would have conferred a substantial adaptive advantage to mammalian ancestors and constitutes a critical step in the evolution of human imagination and language.

References:

Corballis, M. C. (2019). Mental time travel, language, and evolution. *Neuropsychologia*, *134*(August), 107202. https://doi.org/10.1016/j.neuropsychologia.2019.107202

Raichle, M. E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience*, *38*(1), 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

Schacter, D. L., & Addis, D. R. (2020). Memory and Imagination: Perspectives on Constructive Episodic Simulation. In Ed. A. Abraham (Ed.), *The*

Cambridge Handbook of the Imagination (pp. 111–131). Cambridge University Press, Cambridge, England. https://doi.org/10.1017/9781108580298.008

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. *Neuron*, *76*(4), 677–694. https://doi.org/10.1016/j.neuron.2012.11.001

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

The Origin of Typological Suffixing Bias in World Languages

Ines Ferreira[1], Leona Polyanskaya[1], and Mikhail Ordin[2,3]

[*]Corresponding Author: mikhail.ordin@gmail.com
[1]Faculty of Psychology, Universidade de Coimbra, Portugal
[2]Medical Faculty, Universidade de Coimbra, Portugal
[3]Coimbra Institute for Biomedical Imaging, Portugal

The tendency to exploit suffixes more often than prefixes to express grammatical meanings in world's languages was identified a century ago. The underlying causes for this suffixing bias are not clear. Moreover, there is contradictory evidence whether this bias is domain-general or restricted to natural languages. We addressed these questions in a three behavioral and one EEG experiments. The data confirm domain-specificity of the phenomenon, and absence of cognitive bias that could lead to the emergence of typological suffixing bias.

Appending an affix to the word stem is one of the most frequently exploited means to express grammatical meaning (e.g., tense-aspect, number, case, person, interrogation, subordination). Most common affixes are suffixes (appended after the stem) and prefixes (appended before the stem), with infixes and circumfixes being relatively rare phenomena. Linguists have identified a clear preference for suffixing in world languages (Cutler et al. 1985; Greenberg 1957; Hawkins and Gilligan 1988; Sapir 1921). In the World Atlas of Language Structures (WALS), Dryer (2005), there are 8 strongly suffixing languages for 1 strongly prefixing language. This asymmetry is often referred to as a typological suffixing bias. The origin of this bias is still debated.

One potential cause of the typological bias might be an existing cognitive bias, making memorization and/or processing of sequences with variable endings easier than those with variable beginnings. Language might be shaped by domain-general cognitive constraints on memory (Gibson 2000), learning (Hall 1991; Kersten et al. 1998), auditory perception (Blevins 2004; Neath, 1993; Macintosh, 1975). These constraints define domain-general cognitive mechanisms, which act to select those variants of language code that are more easily processed by existing cognitive mechanisms. The selected variants are modified and passed on to the next generations by means of social learning and cultural evolution (Christiansen & Chater, 2001; Lewis et al., 2006; Saygin et al., 2003). If the cognitive bias explains typological bias, we should observe its existence across language and

non-language domains. Alternatively, it can be argued that suffixes can be more easily processed by the language-specific cognitive machinery. For example, the beginning of the word may be more important for lexical access than the end of the word, because the pool of potential word candidates becomes increasingly narrower as more and more segmental information is becomes available (Erdeljac & Mildner, 1999; Marslen-Wilson, 1987; Rodd, 2004). Therefore, left-most segments are most critical for the word activation, and variation at the left edge of the word impedes word recognition, disfavoring variable beginnings of the words. Speech production machinery allows for better preservation of the phonetic contrasts at the beginning of the speech units than at the end, which might also restrict the suffixing bias to language domain. I am going to report 3 behavioural and 1 EEG experiments in an attempt to resolve the origin of the typological suffixing bias.

We used an artificial language learning paradigm (Saffran et al., 1996) to study how adding a prefix or suffix to the recurrent stem-like constituents will interfere with learning and recognition of these constituents by statistical learning mechanisms. In Experiment 1, we ran the study with Spanish monolinguals and Basque-Spanish bilinguals, the latter are more familiar with grammatical prefixes, esp. in verbal paradigms, the former do not have experience with grammatical prefixes. In Experiment 2, we ran the study with German and Slovak participants, both languages are strongly suffixing in regard to inflectional morphology, but both make extensive use of verbal derivational prefixes – yet in German verbal prefixes can be detached and used separately from the stem. We study a reverse effect – the effect of language typology on general suffixing bias, and whether it can extend from inflectional to derivational morphology. In Experiment 3, Arabic speakers, who are used to express most grammatical meanings are expressed by vowel alternations inside the stems rather than appendix affixes, did the experiment with artificial language composed of prefixed and suffixed recurrent words in the familiarization stream. In Experiment 4, we measured EEG signal in Portuguese speakers during familiarization and during learning, to understand whether ERPs are stronger to violations of prefixed or suffixed sequences, or "words" of an artificial language. We explored P3 at the final stages of learning as a neural response to morphological violations and N400 as a response to foils during the test. All experiments were run on linguistic and non-linguistic material to explore across-domain transfer.

The data did not show any evidence for pre-existing cognitive suffixing bias. On the contrary, the typological linguistic features affect how prefixed and suffixed sequences are processed. The effect is weaker or non-existent on non-linguistic material, suggesting that the phenomenon is language-specific rather than defined by general cognitive machinery.

References

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns.* Cambridge, England: Cambridge University Press.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Science, 5*, 82-88.

Cutler, A., Hawkins, J. A., & Gilligan, G. (1985). The suffixing preference: A processing explanation. *Linguistics 23*, 723758.

Dryer, M. S. (2005). Prefixing versus suffixing in inflectional morphology. In M. Haspelmath, M. S.Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures.* Oxford, UK: Oxford University Press.

Erdeljac,V., & Mildner,V.(1999). Temporal structure of spoken-word recognition in Croatian in light of the cohort theory. *Brain and Language, 68*, 95-103.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Greenberg, J. H. (1957). Order of affixing: A study in general linguistics. In J. H. Greenberg (Ed.), *Essays in linguistics* (pp. 86–94). Chicago, IL: University of Chicago Press.

Hall, G. (1991). *Perceptual and associative learning*. New York: Clarendon Press.

Hawkins, J. A., & Gilligan, G. (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua 74*, 219259.

Kersten, A. W., Goldstone, R. L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1437-1458.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*, 447-454.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276-298.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory and Cognition 21*, 689-698.

Rodd, J. M. (2004). When do leotards get their spots? Semantic activation of lexical neighbors in visual word recognition. *Psychonomic Bulletin and Review, 11*, 434-439.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science, 274(5294)*, 1926–1928.

Sapir, E. (1921). *Language.* New York: Harcourt Brace.

Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds: Evidence from aphasia. *Brain, 126*, 928-945.

The relationship between community size and iconicity in sign languages

Shiri Lev-Ari[*1], Rose Stamp[2], Connie de Vos[3], Uiko Yano[4], Victoria Nyst[5], and Karen Emmorey[6]

[*]Corresponding Author: shiri.lev-ari@rhul.ac.uk

[1] Dept. of Psychology, Royal Holloway, University of London, Egham, United Kingdom
[2]Dept. of English Literature and Linguistics, Bar-Ilan University, Ramat Gan, Israel
[3] Dept. of Communication and Cognition, Tilburg University, Tilburg, The Netherlands
[4] Anthropological Studies, The Graduate University for Advanced Studies, SOKENDAI, Hayama, Japan
[5] Centre for Linguistics, Leiden University, Leiden, The Netherlands
[6] School of Speech, Language and Hearing Sciences, San Diego State University, San Diego, CA, USA

Past research shows that larger communities develop more sound symbolic languages to overcome their greater communicative challenges. This study tests whether the results extend to sign languages, as it is debated whether iconicity plays the same role in spoken and signed languages. Participants from five different countries guessed the meaning and rated the iconicity of social and non-social signs from 11 different sign languages. Non-social signs from larger sign languages were rated as more iconic than those from small languages.

Communication is harder in larger communities: Members of larger communities experience greater linguistic variability, more interaction with strangers, and greater bottlenecks for information spread. Larger communities overcome the greater communicative difficulties they encounter by creating languages that are easier to learn and use (Lupyan & Dale, 2010). In particular, languages spoken by more people are more iconic (Lev-Ari et al., 2021), presumably because iconicity facilitates language acquisition and processing (e.g., Imai et al., 2008; Sidhu et al., 2020). We tested whether larger sign languages are also more iconic, as it is debated whether iconicity plays the same role in spoken and signed languages. We also tested whether the effect of community size depends on semantic category.

Non-signers from five different countries (N=178) guessed the meaning and rated the iconicity of signs from five large sign languages (>500,000 signers),

and six small ones (<3,000 signers). The raters lived in different countries to those in which the sign languages are used. Twenty concepts were filmed in each sign language: ten social (e.g., 'friend') and ten non-social (e.g., 'garlic') matched for frequency and concreteness. As predicted, participants rated signs from larger sign languages as more iconic, but this effect was limited to non-social signs (β=-0.43, SE=0.05, z=-8.07, p<0.001). This effect was mainly driven by the fact that signs rated as low in iconicity were almost exclusively from smaller sign languages (See Figure 1). Furthermore, iconicity ratings and guessing accuracy were more aligned in signs from larger sign languages (β=0.12, SE=0.05, t=2.23), potentially because signs from larger languages are less likely to rely on culture-specific iconicity.

Lastly, exploratory analyses revealed an interaction between concreteness and iconicity (β=0.52, SE=0.24, z=2.14, p=0.03) such that iconicity was negatively correlated with concreteness in small languages (β=-0.57, SE=0.16, z=-3.54, p<0.001) but was not associated with concreteness in large sign languages. This finding suggests that, contrary to prior proposals (e.g., Lupyan & Winter, 2018), the prevalence of iconicity is not constrained by the difficulty of representing abstract concepts iconically.

The study supports the claim that social structure shapes language. Specifically, it shows that community size can influence lexical form, leading larger languages to be more iconic. It further shows that iconicity plays a parallel role in spoken and signed languages.

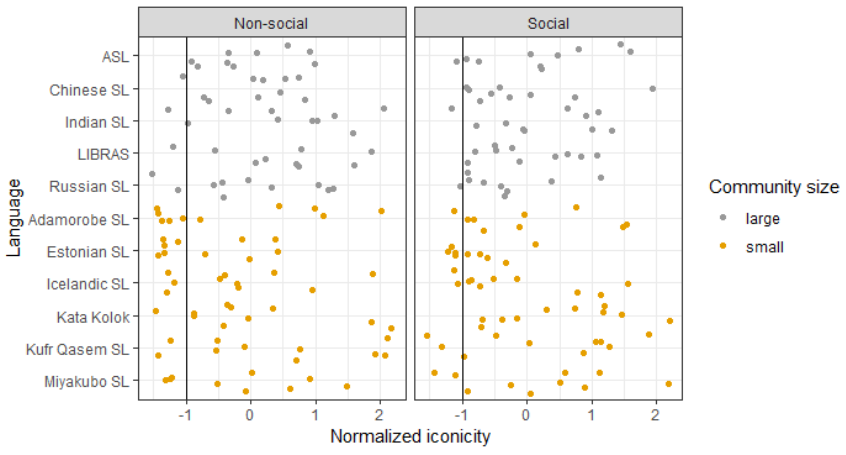


Figure 1. Normalized average iconicity of signs by Language and Semantic Domain. Each dot represents the normalized average iconicity rating for one sign. The y-axis indicates the language. Signs from languages with large communities appear in grey and signs from languages with small communities appear in yellow. Signs to the left of the vertical line received an average iconicity rating that is more than 1 SD below the mean.

References

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*, 54–65.

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*, *51*, e8559.

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic?. *Philos. Trans. R. Soc. B: Biol. Sci., 373, 1752*, 20170137.

Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. *Language and cognition, 12, 1*, 164-181.

Acknowledgements

The study was funded by a grant from the National Institute on Deafness and other Communication Disorders awarded to K.E. (R01 DC010997) and by a Leverhulme grant awarded to S.L (RPG-2022-270). We would like to thank Sara Lanesman for her help in stimuli preparation.

Toward physical chimpanzee vocal tract models

Axel Ekström[1,2,3*] and Rui Tu[4]

[1]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[2]Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden
[3]Centre for Cultural Evolution, Dept. of Psychology, Stockholm University, Stockholm, Sweden
[4]Department of Linguistics and Philology, Uppsala University, Uppsala, Sweden
[*]Corresponding Author: axeleks@kth.se

We argue that physical models and demonstrations can contribute to more informed discussions on the evolution of vocal tract morphology and its phonetic potential. Here, we present the first physical models of non-human vocal tracts, based on chimpanzee vocal tract imaging data. The purpose of this exercise was not performing physical experiments but educational, with the goal of highlighting differences in vocal anatomy.

1. Introduction

Physical vocal tract models are useful for demonstrating the mechanics of speech production, providing a tangible way to understand how sound is generated and modified. Potentially, such modeling allows even audiences unfamiliar with speech acoustics to visualize the shape and movements of the vocal tract components (Arai, 2012). We present physical models designed to emulate vocal tract configurations employed by chimpanzees (*Pan troglodytes*) while producing "*hoo's*" – an apparently vowel-like call acoustically overlapping with close back rounded vowel [u] (Grawunder et al., 2022). We argue that making such models more widely available may educate a broader audience about the significance of inter-species vocal tract physiological differences.

2. Modeling

A subadult chimpanzee vocal tract was traced after Nishimura (2005), and partitioned into equidistant sections after Fant (1992). We modeled a set of four vocal tract shapes, corresponding to (1) the tracing without any additional inference or addition, (2) the tracing with added lip protrusion (inferred from Fant, 1971); (3) the tracing with retracted tongue body (see Takemoto, 2008); and (4) protruded lips and retracted tongue body – i.e., a model combining (2) and (3). Area functions were implemented in the *TubeN* software, which implements the Liljencrants-Fant algorithm to predict resonance frequencies from the area-length

relationships of a vocal tract represented as a concatenated tube sequence (Liljencrants & Fant, 1975). Lip protrusion was conservatively modeled using dimensions from human [u] articulation (Fant, 1971), while tongue body retraction was simulated by narrowing the oropharyngeal segment areas in a stepwise manner. The full modeling considerations are available elsewhere (Ekström et al., 2025a).

3. 3D Printing

The *TubeN* software allows for the semi-automatic transcription of an input tube sequence to 3D-printable material (Zhang et al., 2024). We used a Prusa MK 3.9S 3D printer, and models were printed using polylactic acid (PLA). Figure 1 displays the printed version of the fourth configuration (combined lip protrusion and tongue retraction), which best approximated the acoustic properties of chimpanzee [u]-like vocalizations (Ekström et al., 2025a).

4. Future work

Recent literature has highlighted roles of realistic biomechanics in explicating call behavior in living primates, and any relationships call production may have to speech production (Berthommier, 2020; Ekström, 2024). We suggest that the integration of more general speech production and speech acoustics-based modeling (Fant, 1992) and teaching (Arai, 2012) may facilitate consilience toward this end. A plausible next step would be the incorporation of modeling the chimpanzee nasal tract (Samarat & Matsuzawa, 2016; Bastir, Sanz-Prieto, & Burgos, 2022), the influence of which has been widely recognized (Lieberman, 1984), but never incorporated into modeling efforts (*cf.* Havel, Sundberg, Traser, Burdumy, & Echternach, 2023).

While literature on primate call acoustics and their contextual use is extensive, comparatively little is known about the constraints on vocal production. Through "reverse engineering" paradigms, such as that illustrated here, this may be subject to change. Modeling primate vocal tract physiology may help advance the field of primate vocalization and communication toward more integrative work in the future. For example, an exhaustive collection of physical vocal tract models corresponding to vocal tract estimates for nonhuman primates (Nishimura, 2005; Takemoto, 2008) and human ancestors (Lieberman & McCarthy, 1999; Ekström et al., 2025b) may facilitate the integration of such data into the broader literature on the evolution of vocal production behavior.

Acknowledgements

The results of this work and the tools used will be made more widely accessible through the national infrastructure Språkbanken Tal under funding from the Swedish Research Council (2017–00626).

Additional funding was provided by the Swiss National Science Foundation (PCEFP1_186841).

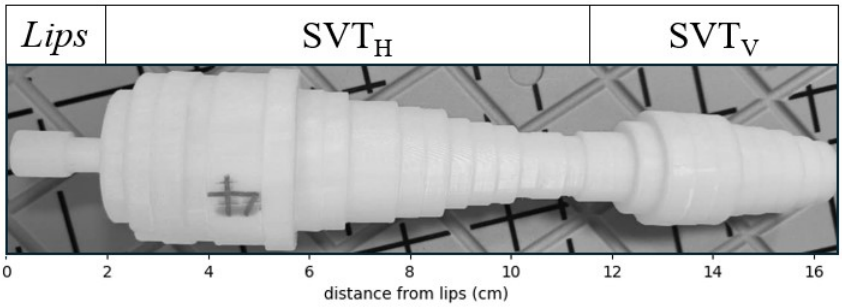


Figure 1. 3D-printed tube chimpanzee vocal tract model, modeled on a tracing after Nishimura (2005). Approximate positions of the lips, and horizontal and vertical sections of the vocal tract (SVT_H, SVT_V) are annotated. SVT_H encompasses the oral cavity and back of the throat; the modeled section of the SVT_v corresponds to the oro- and laryngopharynx. Like all nonhuman primates, the chimpanzee SVT_H is markedly longer than the SVT_v (Lieberman, 1984; Nishimura, 2005).

References

Arai, T. (2012). Education in acoustics and speech science using vocal-tract models. *The Journal of the Acoustical Society of America*, *131*, 2444-2454.

Bastir, M., Sanz-Prieto, D., & Burgos, M. (2022). Three-dimensional form and function of the nasal cavity and nasopharynx in humans and chimpanzees. *The Anatomical Record*, *305*(8), 1962–1973.

Berthommier, F. (2020). Monkey vocal tracts are not so "speech ready". In *Proceedings of 12th international conference on voice physiology and biomechanics* (p. 59). Université Grenoble Alpes.

Ekström, A. G., Bortolato, T., Crockford, C., Lameira, A., Masi, S., Moran, S., Nellissen, L., Shumaker, R. W., Wittig, R. W., & Edlund, J. (2025a, May). *Human vowels and great ape vowel-like calls reflect species-unique constraints on production.* `https://osf.io/preprints/psyarxiv/qru5h_v1`. (PsyArXiv Preprint.)

Ekström, A. G., Gärdenfors, P., Snyder, W. D., Friedrichs, D., McCarthy, R. C., Tsapos, M., Tennie, C., Strait, D. S., Edlund, J., & Moran, S. (2025b). Correlates of vocal tract evolution in late pliocene and pleistocene hominins. *Human Nature*, *36*, 22—69.

Ekström, A. G. (2024). Correcting the record: Phonetic potential of primate vocal tracts and the legacy of Philip Lieberman (1934–2022). *American Journal of Primatology*, e23637.

Fant, G. (1971). *The acoustic theory of speech production: With calculations based on x-ray studies of russian articulations.* Mouton De Gruyter.

Fant, G. (1992). Vocal tract area functions of Swedish vowels and a new three-parameter model. In *Proceedings of the 2nd international conference on*

spoken language processing (ICSLP 1992) (pp. 807–810). ISCA.

Grawunder, S., et al.. (2022). Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. *Philosophical Transactions of the Royal Society B*, *377*, 20200455.

Havel, M., Sundberg, J., Traser, L., Burdumy, M., & Echternach, M. (2023). Effects of nasalization on vocal tract response curve. *Journal of Voice*, *37*, 339–347.

Lieberman, D. E., & McCarthy, R. C. (1999). The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *Journal of Human Evolution*, *36*(5), 487-517.

Lieberman, P. (1984). *The biology and evolution of language.* Harvard University Press.

Liljencrants, J., & Fant, G. (1975). Computer program for VT-resonance frequency calculations. *STL-QPSR*, *16*, 15–21.

Nishimura, T. (2005). Developmental changes in the shape of the supralaryngeal vocal tract in chimpanzees. *American Journal of Physical Anthropology*, *126*(2), 193-204.

Samarat, K., & Matsuzawa, T. (2016). A computational model of the anatomy of realistic chimpanzee nasal airways. *Imperial Journal of Interdisciplinary Research*, *2*(2), 357-364.

Takemoto, H. (2008). Morphological analyses and 3d modeling of the tongue musculature of the chimpanzee (*Pan troglodytes*). *American Journal of Primatology*, *70*(10), 966-975.

Zhang, K., Song, R., Tu, R., Edlund, J., Beskow, J., & Ekström, A. G. (2024). Modeling, synthesis and 3D printing of tube vocal tract models with a codeless graphical user interface. In *Proceedings from FONETIK 2024* (pp. 155–160). Stockholm, Sweden.

Under What Conditions Does Syntactic Alternation Culturally Emerge?

Ibuki IWAMURA[*1] and Takashi HASHIMOTO[1]

[*]Corresponding Author: ibuki_iwamura@jaist.ac.jp
[1]Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Japan

This study investigates the conditions under which syntactic alternation encoding different conceptualizations of the same situation emerges through cultural evolution. We incorporated conceptualization into the Iterated Learning Model, modeling meaning as the specific way an individual construes a situation. Simulations show that alternation emerges only when a learner perfectly infers the speaker's conceptualization. The result suggests that the inferability of conceptualizations determines whether any conceptualized meaning is linguistically encoded through cultural evolution. This study may contribute to understanding how the boundary between syntax and pragmatics arises in language evolution.

1. Introduction

Human language not only denotes objective referents in the external world but also encodes subjective meanings shaped through conceptualization —the process by which individuals perceive and interpret the world. Conceptualization is often reflected in linguistic structures, with syntactic alternation serving as a common means of encoding different ways of conceptualizing the same situation. As Cognitive Grammar (Langacker, 1990) has shown, conceptualization partially determines linguistic structure (e.g., A resembles B vs. B resembles A). Such alternations may not have existed from the origin of language but rather emerged through cultural evolution. A key model of cultural evolution, the Iterated Learning Model (ILM), has demonstrated how language gradually becomes systematic through intergenerational transmission, even if the initial language had no structure (Kirby, 2002; Kirby, Tamariz, Cornish, & Smith, 2015). However, most previous research on ILM has adopted a "situation = meaning" framework instead of "conceptualization of a situation = meaning", making it insufficient for explaining the cultural evolution of syntactic alternation. This study extends ILM by introducing conceptualization to examine the conditions under which syntactic alternations that encode different conceptualizations of the same situation can emerge through cultural evolution.

2. Methodology

We developed a novel ILM that models meaning as "conceptualization of a situation", focusing on binary relations (e.g., A is above B vs. B is below A). We

define syntactic alternation as follows:

- Both forms have a common set of lexical elements.
- Morphosyntactic differences (e.g., word order or morphological marking) correspond to differences in conceptualization.

Thus, we introduced two evaluation metrics (specifically to evaluate the role of word order in morphosyntactic differences):

- Lexical dissimilarity: The mean edit distance between lexical sets.
- Word order dissimilarity: The mean edit distance between word sequences.

Our model modified the Definite Clause Grammar in Kirby (2002) to incorporate conceptualization into semantic representations. In this model, language is represented as a mapping between meanings and forms. Meanings are formalized as $predicate_i(argument_j, argument_{k\neq j})/CV$, where CV (Conceptualization Value) is a binary value (0 or 1) capturing the essence of subjective construal in binary relations (e.g., trajector/landmark selection, active/passive voices). Agents use three rule-based learning algorithms—Chunk, Category-Integration, and Replace—to generalize linguistic rules. Chunk learning specifically generalizes rules with distinguishing CVs, whereas other algorithms learn independently of CVs. This means that rules for entire events incorporate CVs, while rules for the parts of events do not. The production algorithm generates forms in accordance with CVs by combining learned or invented rules, which enable parents to express the entire semantic space. Children receive only half parental productions, simulating a 'bottleneck effect'. Assuming children's reliance on extralinguistic and paralinguistic cues (e.g., gaze and pointing gestures) to infer parental CVs, we investigated the influence of inference accuracy on the emergence of alternation.

3. Results and Discussion

The simulation results revealed that syntactic alternation can emerge only when children can perfectly infer the parent's conceptualization. Even a slight imperfection in the accuracy of inference prevents the emergence of alternation. Specifically, such imperfection hinders the child's correct chunk learning of parental linguistic rules with distinct CVs. Consequently, the child acquires erroneous generalized rules that confuse originally distinct CVs. Low inference accuracy amplifies the child's pressure for simplification within cultural transmission, where the parent's pressure for expressing distinct conceptualizations is also present. Conversely, perfect inference accuracy establishes an equilibrium between these two pressures.

This suggests that information easily inferred from extralinguistic and paralinguistic cues becomes integrated into a syntactic system (e.g., word order or morphosyntactic marking) through cultural evolution, eliminating the need for further

inference. Conversely, information that is difficult to infer from extralinguistic and paralinguistic cues fails to integrate into the syntactic system through cultural evolution and instead remains within the domain of pragmatic inference. Therefore, conceptualizations that are difficult to infer remain implicit and rely on contextual interpretation rather than explicit grammatical coding. Our finding suggests that the ease of inferring information determines whether it integrates into linguistic structure through cultural evolution. This study may contribute to understanding how the boundary between syntax and pragmatics arises in language evolution.

4. Conclusion

By integrating conceptualization into ILM, this study demonstrates that syntactic alternation encoding subjective meanings can evolve culturally if children can reliably infer the conceptualization behind linguistic expressions. However, the current model focuses solely on intergenerational transmission, neglecting the effects of social interaction and communication. Future research could explore how interlocutors' mutual understanding influences the grammaticalization of conceptual distinctions.

Our findings advance the discussion of why human language encodes conceptualization beyond objective referents and provide insights into the role of cultural evolution in shaping linguistic structure.

Acknowledgements

This work was supported by JST BOOST, Japan Grant Number JPMJBS2425.

References

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. *Linguistic evolution through language acquisition: Formal and computational models*, 173–204.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Langacker, R. W. (1990). *Concept, image, and symbol: The cognitive basis of grammar.* Mouton de Gruyter.

Using predictive dynamics as a window into the evolution of Language and Music

Felix Haiduk[*1] , Nicholas Harley[2], John McBride[3], and Tudor Popescu[4]

[*]Corresponding Author: haiduk.felix@gmail.com
[1]Acoustics Research Institute, Austrian Academy of Science, Vienna, Austria
[2]Artificial Intelligence Lab, Vrije Universiteit Brussel, Belgium
[3]University of Vienna, Austria
[4]Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Austria

We suggest the application of a framework from Evolutionary Systems Biology to language and music evolution to complement the multi-component approach to music and language with a synthesis approach. Our future goal is to reconstruct the order of trait evolution and their impact for transitions between protomusic and protolanguage stages. To this end, we focus on predictive dynamics in vocalisation systems, operationalising them using an information-theoretic model. We present first results, employing different information-theoretical analyses to a cross-cultural song-speech dataset.

1. Background

Humans are unique among social animals in having both music and language. A key approach to studying their evolution is identifying shared traits, like vocal learning or entrainment, across species and mapping them onto phylogenetic trees (Honing et al., 2015). Related to these traits, cross-cultural research revealed near-universal features specific for or shared by music and language (Savage et al., 2015). However, many crucial traits emerged after our split from chimpanzees, limiting this method's explanatory power. Understanding the evolutionary sequence and interaction of these traits could refine hypotheses about protolanguage and protomusic systems and their evolutionary relation.

2. Aims

We propose adapting an evolutionary systems biology (ESB) framework aiming to identify causal interactions in evolving systems and define their possible evolutionary trajectories. Using dynamical systems modeling, this approach provides mechanistic explanations for phenotypic development and evolutionary transitions. The key domain to assess evolvability and robustness are the component dynamics of the system - in the original ESB framework gene expression dynamics linking genotype to phenotype (Jaeger & Monk, 2021). In transferring this approach, we suggest investigating the domain of predictive dynamics mediating between cognitive traits ("genotypic") and spectrotemporal features ("phenotypic"). We reason that (1) each neurobiologically implemented cognitive capacity enables new predictive dynamics, (2) these dynamics shape

the acoustic features that actualise prediction in social interactions, and (3) they can be quantified using computational and dynamical systems modeling. Measuring and simulating these predictive dynamics would allow us to analyse evolvability and characterise evolutionary transitions in a causal-mechanistic way. Integrating this approach with information-theoretic modeling of cognitive predictive processes in vocalisation systems could thus help reconstruct the evolutionary pathways of music and language.

3. Methods

In the current study we approached a future implementation of the ESB framework by investigating predictability in song, instrumental music, and recited and descriptive speech from a cross-cultural dataset, comprising examples from 21 language families/cultural backgrounds from all continents (Ozaki et al., 2024). We conceptualise predictive dynamics as interface between song/speech acoustics and cognitive capacities/goals and quantified it using both n-gram analysis and the information-theoretic model IDyoMS (Pearce, 2005). This model operationalises auditory perception through "viewpoints" like pitch and duration. We computed information content and entropy across viewpoints and examined the resulting networks with graph-theoretic measures.

4. Results

We found that mean information content increased from descriptive to recited speech to song and instrumental music, forming a continuum with the highest predictability in instrumental music. However, we also observed that both network topographies and observed predictive patterns were similar between conditions, suggesting that construction and choice of viewpoints is highly influential. We also found song and speech to differ in specific subnetworks, suggesting that the same set of perceptual dimensions may still lead to different predictive dynamics.

5. Discussion and conclusion

We propose that a framework from ESB can be applied to cognition to gain empirical insight into music and language evolution. We will discuss current limitations and future directions of operationalising predictive dynamics using information-theoretic viewpoint networks to investigate the evolutionary differentiation of song and speech.

References

Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140088–20140088. https://doi.org/10.1098/rstb.2014.0088

Jaeger, J., & Monk, N. (2021). Dynamical modules in metabolism, cell and developmental biology. *Interface Focus*, *11*(3). https://doi.org/10.1098/rsfs.2021.0011

Ozaki, Y., Tierney, A., Pfordresher, P. Q., McBride, J. M., Benetos, E., Proutskova, P., Chiba, G., Liu, F., Jacoby, N., Purdy, S. C., Opondo, P., Fitch, W. T., Hegde, S., Rocamora, M., Thorne, R., Nweke, F., Sadaphal, D. P., Sadaphal, P. M., Hadavi, S., … Savage, P. E. (2024). Globally, songs and instrumental melodies are slower and higher and use more stable pitches than speech: A Registered Report. *Science Advances*, *10*(20). https://doi.org/10.1126/sciadv.adm9797

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition.* [City University of London, UK]. http://openaccess.city.ac.uk/1189/

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 Sciences*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Using Neural Networks for part-of-speech analysis to investigate syntactic evolution and change

Carter Smith[1] and Frederik Hartmann[1]

[1]Department of Linguistics, University Of North Texas, Denton, Texas, USA

Using computational models to study syntactic evolution necessitates a clear understanding of syntax in historical low-resource languages. One avenue to gain insights into syntactic processes is employing Neural Network Part-of-Speech analyzers. This study investigates using Bi-directional Long-Short Term Memory models to tag POS in 5 historical Germanic languages. First, we employ mono-language tagging models, then utilize ablation to investigate class-wise predictive patterns between languages and observed syntactic patterns. Finally, we test multilingual models that utilize cross-linguistic patterns for prediction.

1. Introduction

This study investigates the value of the application of deep learning techniques to Part-of-Speech (POS) tagging, which here refers to the assignment of a main word class/syntactic category label to each word in a sequence, in historical Germanic for research in language evolution. (For a discussion of POS tagging as compared to full morphological tagging, see Scherrer (2021)). Chiche and Yitagesu (2022) describes the various efforts and techniques made in automatic POS annotation, however, the application of these methods to historical languages is limited. As a test case, the models are trained on five historical low-resource languages: Old English, Old Saxon, Old Icelandic, Old High German, and Gothic. We first test mono-language models and then analyze how syntactic patterns impact class-wise predictability in the languages. We then investigate the trained embedding space and predictive accuracies across all POS to infer diachronic relationships across the languages.

2. Data and Methods

2.1. *Data*

The data used in this study were extracted from the following corpora: Old Saxon from the HeliPaD: a parsed corpus of Old Saxon (Walkden, 2016), Old English from the York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE) (Taylor, Warner, Pintzuk, & Beths, 2003), Old Icelandic from the Icelandic Parsed

Historical Corpus or IcePaHC (Wallenberg, Ingason, Sigurðsson, & Rögnvaldsson, 2011), Old High German from the Referenzkorpus Altdeutsch (Zeige et al., 2025), and Gothic from Project Wulfila (De Herdt et al., 1997).

2.2. *Models*

The first set of models constructed for this study utilizes word-level embeddings, which are trained alongside the models to extract as much information about individual words as possible (So as to use syntactic-level information as the primary factor in inferring POS). This is followed by a series of Bi-LSTM layers to make context-sensitive predictions about POS, thus drawing on the syntactic frame to identify POS. We further train embeddings directly on the syntactic categories with skip-gram models to analyze the relationship between parts of speech in the different languages.

To eliminate the confounding factor of dataset size when making comparisons between the respective languages, we conducted an ablation study. This allows us to observe more directly which POS are more or less predictable between the tested languages. This way, we can compare these predictions with differences in the syntactic patterns between languages to identify where accuracies are markers of less stable linguistic states and can point to mechanisms of language evolution.

Table 1. Accuracy of mono-language Bi-LSTM Models Overall and by POS.

tag	OE	OS	OHG	OI	GO
Overall	91.34%	87.68%	80.69%	89.34%	89.43%
N	93.68%	88.50%	76.66%	90.14%	88.68%
VB	92.48%	82.18%	76.37%	85.95%	86.90%
AP	89.01%	96.58%	92.66%	88.14%	96.87%
PRO	94.69%	95.74%	93.73%	97.01%	95.63%
D	84.57%	87.33%	91.17%	82.20%	-
ADV	84.76%	88.29%	77.63%	93.41%	85.68%
CONJ	94.03%	93.66%	76.91%	99.13%	93.88%
ADJ	83.31%	66.28%	52.13%	64.16%	74.66%
C	97.16%	72.72%	-	97.80%	-
Q	95.59%	93.97%	-	92.32%	-
NUM	95.38%	70.83%	40.22%	85.53%	20.37%
NEG	99.87%	96.92%	-	97.97%	-
PART	65.78%	37.50%	95.40%	48.25%	0.00%
WH	90.62%	77.77%	79.37%	76.76%	-
INTJ	90.94%	0.00%	33.33%	0.00%	0.00%

3. Results and Discussion

Preliminary results have shown a high degree of accuracy on the predictions for the monolingual models, as shown in **Table 1**

In our skip-gram, we have identified notable differences in the relative distances between POS between languages, with, for example, conjunctions and complementizers patterning much more closely together in Old Saxon when compared to Old English.

Additionally, comparisons of class-wise predictive accuracies reflect observed syntactic patterns. For example, the corpus analysis in Bech et al. (2024) showed that the order of adjective and noun is far more restricted in OE when compared to OHG, OI, and OS. This is reflected in the model's lower predictive accuracy of ADJ in these languages, which, in turn, indicates that this pattern is in a less stable and learnable state.

Furthermore, the ablation study allows us to identify more notable patterns. For example, the OHG determiner shows a pattern of learnability distinct from the other languages. Rather than reaching peak predictive accuracy and leveling off at a fairly low sample size (as is common for function words), predictive accuracy continues to grow with sample size (resembling more broadly the pattern of learnability of content words). Ablation has also allowed us to note other impacts on predictability, like the greater dialect diversity in OHG on the predictability of certain POS (like conjunctions) when compared to the other languages.

4. Further Modeling

Following this analyses, we seek to develop models that adaptively utilize information at other levels (such as at the character level, to encode morphological information), and work using cross-lingual datasets, while employing similar methods of quanititative-based qualitative analysis, as well as test full morphosyntactic tagging schemes to gain further insight into diachronic patterns.

References

Bech, K., Booth, H., Börjars, K., Breban, T., Petrova, S., & Walkden, G. (2024). Chapter 3: Noun phrase modifiers in early germanic: A comparative corpus study of old english, old high german, old icelandic, and old saxon. In K. Bech & A. Pfaff (Eds.), *Noun phrases in early germanic languages* (p. 71-141). Language Science Press.

Chiche, A., & Yitagesu, B. (2022). Part of speech tagging: A systematic review of deep learning and machine learning approaches. *Journal of Big Data*, *9*(1).

De Herdt, T., Kinnaer, F., Landau, D., Tannert, R., Van Assche, S., & Van Loon, J. (1997). *Project wulfila: an online edition of the gothic bible.* (University of Antwerp)

Scherrer, Y. (2021). Adaptation of morphosyntactic taggers: Cross-lectal and multilectal approaches. *Similar Languages, Varieties, and Dialects: A Computational Perspective*, 138.

Taylor, A. D., Warner, A. R., Pintzuk, S., & Beths, F. (2003). *The york-toronto-helsinki parsed corpus of old english prose (ycoe).* (Electronic texts and manuals available from the Oxford Text Archive)

Walkden, G. (2016). The helipad. *International Journal of Corpus Linguistics*, *21*(4), 559-571.

Wallenberg, J. C., Ingason, A. K., Sigurðsson, E. F., & Rögnvaldsson, E. (2011). *Icelandic parsed historical corpus (icepahc).* (Version 0.9)

Zeige, L. E., Schnelle, G., Klotz, M., Donhauser, K., Gippert, J., & Lühr, R. (2025). *Deutsch diachron digital - referenzkorpus altdeutsch (version 1.2).* (Humboldt-Universität zu Berlin)

We do not frequently borrow words that we do not borrow frequently: the search for quantifiable characteristics of swadeshness

Ilia Afanasev[*1,2]

[*]Corresponding Author: ilia.afanasev.1997@gmail.com
[1]Department of Slavonic Studies, University of Vienna, Vienna, Austria
[2]Columbia School of Linguistics Society

The talk explores the criteria a historical linguist can use to build a vocabulary list. The main hypothesis states that these criteria are quantifiable and automatically detectable. To test the hypothesis, the study employs the East Slavic language material and the shortened version of existing word lists. The experiments provide supporting evidence for the hypothesis, and the results call for further study, especially on choosing more effective automatic detection methods.

The talk examines the concept of swadeshness (Dellert & Buch, 2016) on the material of East Slavic lects[1] from a formal standpoint, not directly tackling the controversial topic of lexicostatistics (Campbell & Poser, 2008), but rather discussing the ways to formalise this concept in a language-aware way (Borin, 2012; Prokić & Moran, 2013). One of the purposes of such examination is to complement the notion of low borrowability with the set of bottom-up corpus-driven criteria (Divjak, Sharoff, & Erjavec, 2017). The study presents three main hypotheses, one entailing the other.

H1. It is possible to define swadeshness, a parameter that facilitates the creation of Swadesh-like lists[2], as a combination of the following criteria: historical stability, relatively high frequency, relatively low degree of colexification, clear monosemy[3], and stylistic neutrality.

H2. The main characteristics of swadeshness, given in **H1**, are quantifiable and thus detectable in any kind of sequential data, whether word lists or corpora.

[1]The study uses a neutral term *lect* to denote any variety: idiolect, dialect, sociolect, or standard. The purpose of this term is to eliminate the possible appearance of a language / dialect hierarchy that significantly complicates any kind of linguistic study (Otheguy & Stern, 2011)

[2]A small (up to 400) list of meanings that are denoted in each lect, according to the author of the original work (Swadesh, 1955), using the words that remain unchanged in a lect for the longest period of time. This talk employs the term *Swadesh-like*, because there are at least three highly used variations of this list (Holman et al., 2008; Tadmor, Haspelmath, & Taylor, 2010; Kassian, Starostin, & Dybo, 2010) that do not match between themselves or with the original, due to the differences in the methods to detect the (lack of) borrowability.

[3]Or, in terms of monosemy detection-oriented frameworks, easy detection of the lexical meaning (Sabar, 2018).

H3. It is thus possible to check how much the words included in the existing Swadesh-like lists differ from the other parts of the vocabulary in terms of their usage by the speakers.

Previous studies on the topic (Borin, 2012; Kassian et al., 2010; Burlak, 2021) have already investigated some of the criteria mentioned in **H1**, their quantifiability (Dellert & Buch, 2016), and automatic (Afanasev, 2023) detection (**H2**). But, due to the research design, adapted to solve other equally significant issues, they did not fully explore the interconnection between these problems. The presented study, on the contrary, accentuates the swadeshness being the set of quantifiable characteristics of the lexical item (**H3**).

Unlike previous studies that conducted the research on a wide scale, in terms of both lects (Dellert & Buch, 2016) and the number of items under study (Afanasev, 2023), this work reduces the scope. The material is limited to the corpora of two modern standard lects, Ukrainian and Belarusian, and one historical lect, Ruthenian (the last common ancestor of these standards). For Ukrainian, the study presents a new corpus, based on UA-GEC (Syvokon, Nahorna, Kuchmiichuk, & Osidach, 2023), which is reusable for other linguistic studies. The research utilises the corpus material because the criteria, stated in **H1**, require its presence for the bottom-up linguistic study. In addition, this talk takes only one subset from the original Swadesh 100-item word list, the concepts PERSON, MAN, WOMAN (the most general words for people), DIE, and KILL (the words quite frequent in the legal texts (Afanasev, 2023), which is crucial for the study involving historical corpora). To support the analysis, the research uses a group of words that should not match the criteria of swadeshness, being frequently borrowed (Haspelmath & Tadmor, 2009), but still either denote the people or their actions (ENEMY, COMMAND OR ORDER, RULE OR GOVERN) or represent the legal terminology(CITIZEN, WITNESS). The narrow scope of the material, despite severely limiting the generalisability of the results, provides the basis for a more effective qualitative interpretation.

The study presents the attempt at formalising the presented criteria (for instance, treating differences in *ipm* (item per million in corpus) frequency as the rate of change for the word presence in the lect on the (micro-)diachronic scale). It also discusses the automatic means to test specific words against these criteria, including transformer-based neural networks (Vaswani et al., 2017). The first stage of the research is thus gathering the required information for the words in Ukrainian, Belarusian and Ruthenian. The second stage is training a transformer-based neural network for the detection of Swadesh-like list items on the material of Ukrainian and Belarusian corpora. The third stage includes the evaluation of the network on the material of Ruthenian, using the gathered information about the detected items to check their swadeshness. For cross-evaluation (Prokić & Moran, 2013; Afanasev, 2024), the study uses information-based criteria, previously employed for word list data (Dellert & Buch, 2016).

The study demonstrates that while some of the **H1** swadeshness criteria, such as frequency, are formalised, easily quantifiable, and automatically detectable, the other, for instance, clear monosemy of meaning, may present difficulties for formalisation (**H1**) or quantification (**H2**). As for automatic detection (**H2**), especially in corpora, the talk underscores the need for further research into the distribution of these words to choose the most effective algorithm. The talk concludes with an analysis of differences by the given criteria between Swadesh-like list items and other parts of the vocabulary in order to demonstrate that checking cross-linguistic borrowability is not the only way to detect Swadesh-like list items (**H3**).

Acknowledgements

I would like to thank anonymous reviewers for their comments that greatly helped me to improve the abstract. I am also grateful to O. Lyashevskaya, G. Starostin, and E. Kelih for the insightful discussion on the nature of swadeshness.

References

Afanasev, I. (2023). Multi-lect automatic detection of Swadesh list items from raw corpus data in East Slavic languages. In N. Tahmasebi, S. Montariol, H. Dubossarsky, A. Kutuzov, S. Hengchen, D. Alfter, F. Periti, & P. Cassotti (Eds.), *Proceedings of the 4th workshop on computational approaches to historical language change* (pp. 76–86). Singapore: Association for Computational Linguistics.

Afanasev, I. (2024). The Cross-Evaluation Crux for Computational Phylogenetic Linguistics. In M. Bakaev, R. Bolgov, A. V. Chugunov, R. Pereira, E. R, & W. Zhang (Eds.), *Digital Geography* (pp. 75–89). Cham: Springer Nature Switzerland.

Borin, L. (2012). Core Vocabulary: A Useful But Mystical Concept in Some Kinds of Linguistics. In D. Santos, W. Lindén, & W. Ng'ang'a (Eds.), *Shall We Play the Festschrift Game?* (pp. 53–65). Berlin, Heidelberg: Springer Berlin Heidelberg.

Burlak, S. (2021). Stability and frequency: is there a correlation? *Journal of Language Relationship*, *19*(3-4), 293–307.

Campbell, L., & Poser, W. J. (2008). *Language Classification: History and Method.* Cambridge: Camridge University Press.

Dellert, J., & Buch, A. (2016). Using computational criteria to extract large Swadesh Lists for lexicostatistics. In C. Bentz, G. Jäger, & I. Yanovich (Eds.), *Capturing phylogenetic algorithms for linguistics.* Tübingen.

Divjak, D., Sharoff, S., & Erjavec, T. (2017). Slavic Corpus and Computational Linguistics. *Journal of Slavic Linguistics*, *25*(2), 171–98.

Haspelmath, M., & Tadmor, U. (Eds.). (2009). *Loanwords in the World's Languages.* De Gruyter Mouton.

Holman, E., Wichmann, S., Brown, C., Velupillai, C., Müller, A., & Bakker, D. (2008). Explorations in automated language classification. *Folia Linguistica*, *42*, 331–354.

Kassian, A. S., Starostin, G., & Dybo, A. V. (2010). The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship*, *4*, 46–89.

Otheguy, R., & Stern, N. (2011). On so-called Spanglish. *International Journal of Bilingualism*, *15*(1), 85–100.

Prokić, J., & Moran, S. (2013). Black box approaches to genealogical classification and their shortcomings. In L. Borin & A. Saxena (Eds.), *Approaches to Measuring Linguistic Differences* (pp. 429–446). Boston/Berlin: Walter De Gruyter GmbH.

Sabar, N. (2018). *Lexical meaning as a testable hypothesis.* John Benjamins.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, *21*(2), 121–137.

Syvokon, O., Nahorna, O., Kuchmiichuk, P., & Osidach, N. (2023). UA-GEC: Grammatical error correction and fluency corpus for the Ukrainian language. In M. Romanyshyn (Ed.), *Proceedings of the second ukrainian natural language processing workshop (unlp)* (pp. 96–102). Dubrovnik, Croatia: Association for Computational Linguistics.

Tadmor, U., Haspelmath, M., & Taylor, B. (2010). Borrowability and the notion of basic vocabulary. *Diachronica*, *27*(2), 226-246.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in neural information processing systems*, *30*, 1–11.

What exactly is aligned in the case of alignment in whole-body communication? Case studies of pantomimic enactments of short stories.

Klaudia Karkowska [*1], Marta Sibierska [1], Antoni Żyndul [1], and Sławomir Wacewicz [1]

[*]Corresponding Author: karkowska@doktorant.umk.pl
[1] Centre for Language Evolution Studies (CLES),
Nicolaus Copernicus University, Toruń, Poland

Alignment can concern different dimensions, including form. It has been extensively studied in linguistic communication, but there is still a lot to be said about alignment in non-linguistic communication. One research gap is generalising findings about alignment in manual (co-speech) gesture onto whole-body expression (pantomime). Based on case studies of pantomimic enactments of short stories, we show that alignment in form in gesture and pantomime significantly differs, and that the available tools are suitable for the former and not for the latter.

1. Introduction

Alignment is a fundamental mechanism in interaction, whereby we repeat each other's behaviours, so as to communicate successfully (Pickering & Garrod, 2004). Alignment can concern different dimensions of interaction: the tendency to use the same words or the same grammatical structures (e.g., lexical and syntactic alignment in Pickering & Garrod, 2006), also across languages (e.g., Hartsuiker, Pickering & Veltkamp, 2004) or the same cognitive framework, as in the case of spatial referencing (e.g., Levinson, 2003), or even information states (Pickering & Garrod, 2006). In spite of a lot of research on alignment in linguistic communication, there is still a lot to be said about alignment in non-linguistic communication.

One example of a research gap is generalising findings about alignment in manual (including co-speech) gesture onto whole-body expression, in particular pantomime. Advocating for a framework to study multimodal interactions, gesture included, Rasenberg, Özyürek, and Dingemanse (2020) proposed to look at alignment along five dimensions: time, sequence, meaning, form, and modality. As far as alignment in "form" is concerned, it can be dissected into, for instance, alignment in terms of the mode of representation and the shape of

the representation—including handedness, handshape, orientation, or position (Bergmann & Kopp, 2012; Chui, 2014). According to the existing literature on alignment in manual gesture, interactants tend to align both on the level of the mode and shape (e.g., Oben and Brône, 2016). We believe this does not necessarily apply to silent whole-body communication (cf. "pantomime" *sensu* Żywiczyński et al., 2018; cf. with manual, co-speech, or silent gesture, e.g., Rasenberg et al., 2022; Motamedi et al., 2019; Nölle et al., 2018).

2. Study

To elaborate on this assumption, we conducted a series of case studies: we analysed a set of 64 video recordings of a charades-like game, where the participants were arranged into dyads and asked to take turns miming short stories and matching them to sequences of images (the data originally collected for the study reported in Sibierska et al., 2023). We annotated the modes used by the participants in individual rounds of interaction, compared them within dyads, and supplemented them with ratings of similarity in terms of shape (2 levels: "not similar"/"similar").

3. Results and discussion

The annotations and the ratings show that although alignment in the mode of representation is frequent, the specific shapes of representations (i.e. spatial configurations of articulators) remain highly idiosyncratic throughout the interaction. There are some similarities in shape-based associations, but they do not correspond to the understanding of "shape" in the available research on alignment (e.g., Bergmann & Kopp, 2012). This finding can be explained by: (1) the distinctive character of pantomime, where there is one dominant mode—enactment (Müller, 2014)—and other modes only have complementary functions; (2) the fact that pantomime, by definition, is ad hoc and spontaneous, which leads to higher idiosyncrasy. Based on these case studies, we argue that achieving alignment—and, by extension, communicative success (cf. Pickering & Garrod, 2006)—might differ in whole-body communication with respect to other interactions. We also use this opportunity to stress that the treatment of "form" in studies on alignment in non-verbal communication is tailored to manual gesture rather than whole-body pantomime, which is visible in particular in their operationalisation of "shape".

Acknowledgements

This research has been funded by the National Science Centre of Poland under the agreement UMO 2021/43/D/HS2/01866.

References

Bergmann, K., Kopp, S. (2012). Gestural alignment in natural dialogue. In: N. Miyake, D. Peebles, R. P. Cooper (Eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, pp. 1326–1331.

Chui, K. (2014). Mimicked gestures and the joint construction of meaning in conversation. *Journal of Pragmatics* 70, pp. 68–85. https://doi.org/10.1016/j.pragma.2014.06.005.

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. https://doi.org/10.1016/j.cognition.2019.05.001.

Müller, C. (2014). Gestural modes of representation as techniques of depiction. In: C. Müller et al. (Eds.) *Body–Language Communication: An International Handbook on Multimodality in Human Interaction*. De Gruyer Mouton, pp. 1687–1701. https://doi.org/10.1515/9783110302028.1687.

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, pp. 93-104. https://doi.org/10.1016/j.cognition.2018.08.014.

Oben, B., Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics* 104, pp. 32–51. https://doi.org/10.1016/j.pragma.2016.07.002.

Pickering, M. J., Garrod, S. (2006). Alignment as the Basis for Successful Communication. *Research on Language and Computation* 4, pp. 203–228. https://doi.org/10.1007/s11168-006-9004-0.

Pickering, M. J., Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences* 27(2), pp. 169–190. https://doi.org/10.1017/S0140525X04000056.

Rasenberg, M., Özyürek, A., Bögels, S., Dingemanse, M. (2022). The Primacy of Multimodal Alignment in Converging on Shared Symbols for Novel Referents. *Discourse Processes* 59(3), pp. 209–236. https://doi.org/10.1080/0163853X.2021.1992235.

Rasenberg, M., Özyürek, A., Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science* 44, e12911. https://doi.org/10.1111/cogs.12911.

Sibierska, M. et al. (2023). Constraints on communicating the order of events in stories through pantomime. *Journal of Language Evolution* 8(1), pp. 18–32. https://doi.org/10.1093/jole/lzad008.

Żywiczyński, P., Wacewicz, S., Sibierska, M. (2018). Defining pantomime for language evolution research. *Topoi* 37, pp. 307–318. https://doi.org/10.1007/s11245-016-9425-9.

Word representation in the dog brain

Marianna Boros[*1], Lilla Magyari[1,2,3], Dorottya Sz. Rácz[1,4] and Attila Andics[1,5]

[*]Corresponding Author: marianna.cs.boros@gmail.com

[1]Neuroethology of Communication Lab, Department of Ethology, Eötvös Loránd University, Budapest, Hungary

[2]Norwegian Reading Centre for Reading Education and Research, Faculty of Arts and Education, University of Stavanger, Stavanger, Norway

[3]Department of Social Studies, Faculty of Social Sciences, University of Stavanger, Stavanger, Norway

[4]Doctoral School of Biology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary

[5]ELTE NAP Canine Brain Research Group, Budapest, Hungary

Recent neuroscientific reports on dogs' word learning raise the question whether a non-human mind can represent meaning semantically and referentially. I will present converging evidence from non-invasive, awake EEG and fMRI from our lab of object word-elicited semantic expectations in dogs, action words evoking embodied meaning processing, and a representational geometry organized along semantic similarity. Considering the human-analogous neural mechanisms of dogs' word processing, the most parsimonious explanation appears to be one that attributes symbolic representational capacities to dogs.

The human and dog lineages split ~100 mya, but during the ~20,000 years of their domestication dogs have been extensively selected for communication abilities with humans (Hare and Tomasello, 2005; Miklósi and Topál, 2013). Nowadays, more than ever, dogs live immersed in the human socio-linguistic environment, being exposed to speech on a daily basis. Due to their cooperativity and trainability, dogs provide an unparalleled case to study brain activity in an awake, attentive, unrestrained non-human mammal (Bunford et al., 2017). Studying speech processing in dogs thus allows us to track how environmental and genetic factors may shape a mammalian brain during language evolution.

With the recent surge of studies on dogs' speech processing (Cuaya et al., 2022; Mallikarjun et al., 2023, 2021, 2019) and word learning abilities (Bastos et al., 2024b, 2024a; Dror et al., 2022; Fugazza et al., 2021; Fugazza and Miklósi,

2020; Griebel and Oller, 2012; Higaki et al., 2025; Kaminski et al., 2004; Kőszegi et al., 2023; Pilley and Reid, 2011; Ramos and Mills, 2019), the debate about the nature of word representations in non-human animals has resurfaced, raising the question of whether a non-human mind can represent meaning in a truly semantic and referential manner (Markman and Abelev, 2004). However, considering the human-analogous neural mechanisms reported for dogs' word processing (Andics et al., 2016; Boros, Magyari et al., 2024, 2021; Gábor et al., 2020; Magyari et al., 2020), the most parsimonious explanation appears to be one that attributes symbolic representational capacities to dogs.

In my talk, I will present recent converging evidence from a series of non-invasive, awake dog EEG and fMRI experiments from our lab that support the above claim, demonstrating the capacity of dog brains to represent both the auditory forms of words and their associated meanings. I will also highlight differences and critical gaps in knowledge.

First, I will show evidence from a combined fMRI-EEG study, that adult dogs employ sophisticated computational mechanisms to segment speech. We probed dogs' statistical learning capacities for word segmentation by exposing them to speech streams with various distributional cues to word boundaries (Aslin et al., 1998). Their ERPs showed that, like human infants, dogs extract words from continuous speech by computing transitional probabilities – an ability not yet observed in any other non-human mammal. Additionally, fMRI results showed that both domain-general and modality-specific brain regions are involved in syllable sequence processing in dogs, similarly as in humans (Boros, Magyari et al., 2021).

Next, I will demonstrate that auditory word form representations in dogs are coarser-grained than in humans, suggesting that dogs may not attend to each individual speech sound within a word, possibly due to their limited vocabulary. I will present evidence from an fMRI study on action instruction word processing showing that the dog brain does not distinguish between known words and pseudowords created by altering a single speech sound. This lack of sensitivity to phonetic detail during lexical processing resembles patterns observed in human infants before 20 months of age.

Finally, I will present EEG and fMRI findings indicating that these auditory word forms are linked to human-analogous meaning representations. In an EEG experiment we tested dogs' object word understanding in a semantic violation paradigm. We found that a mismatch between prime word and target object evoked a human N400-like ERP effect, revealing that object words can evoke mental representations of the referred objects in dogs, suggesting that dogs' object word understanding is thus, similarly to humans' is referential in nature(Boros, Magyari et al., 2024).

Using fMRI, we also examined dogs' understanding of instruction words for actions. We found stronger activations for instruction words than for nonwords in motor and motor control regions of the dog brain, suggesting that dogs processed the meanings of these words. Semantic effects for action-related words are consistently found in sensorimotor regions in human as well, supporting embodied meaning representations. Additionally, in the dog auditory cortex, activation patterns for words referring to actions requiring locomotion were more similar to each other than to words referring to actions not involving locomotion. This suggests that, as in humans, the auditory cortex in dogs encodes meaning along dimensions of semantic similarity.

Overall, these findings emphasize the relevance of dog-human comparisons for understanding the evolution of language-related abilities and point to dogs as promising models for investigating word processing and symbolic capacities in non-human species. The unexpected parallels in the cognitive and neural mechanisms of word processing in dogs and humans also argue for interpreting these capacities within a symbolic framework.

Acknowledgements

This project was funded by the Hungarian Academy of Sciences, a grant to the MTA-ELTE "Lendület" Neuroethology of Communication Research Group (LP2017-13/2017), the Eötvös Loránd University, and the European Research Council under the European Union's Horizon 2020 research and innovation program (grant number 950159). MB was supported by the ÚNKP-22-4 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund (ÚNKP-22-4-II-ELTE-963). D.S.R. was supported by the EKÖP-24 University Excellence Scholarship Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund (EKÖP-24-3-I-ELTE-58). AA was supported by the National Brain Programme 3.0 of the Hungarian Academy of Sciences (NAP2022-I-3/2022). We thank all dog owners and dogs for their participation in the experiments.

References

Andics, A., Gábor, A., Gácsi, M., Faragó, T., Szabó, D., Miklósi, 2016. Neural mechanisms for lexical processing in dogs. Science (80-.). 353, 1030–1032. https://doi.org/10.1126/science.aaf3777

Aslin, R.N., Saffran, J.R., Newport, E.L., 1998. Computation of Conditional Probability Statistics by 8-Month-Old Infants. Psychol. Sci. 9, 321–324.

https://doi.org/10.1111/1467-9280.00063
Bastos, A.P.M., Evenson, A., Wood, P.M., Houghton, Z.N., Naranjo, L., Smith, G.E., Cairo-Evans, A., Korpos, L., Terwilliger, J., Raghunath, S., Paul, C., Hou, H., Rossano, F., 2024a. How do soundboard-trained dogs respond to human button presses? An investigation into word comprehension. PLoS One 19, e0307189.
Bastos, A.P.M., Houghton, Z.N., Naranjo, L., Rossano, F., 2024b. Soundboard-trained dogs produce non-accidental, non-random and non-imitative two-button combinations. Sci. Rep. 14, 28771. https://doi.org/10.1038/s41598-024-79517-6
Boros, M., Magyari, L., Morvai, B., Hernández-Pérez, R., Dror, S., Andics, A., 2024. Neural evidence for referential understanding of object words in dogs. Curr. Biol. 34, 1750-1754.e4. https://doi.org/10.1016/j.cub.2024.02.029
Boros, M., Magyari, L., Török, D., Bozsik, A., Deme, A., Andics, A., 2021. Neural processes underlying statistical learning for speech segmentation in dogs. Curr. Biol. 31, 5512-5521.e5. https://doi.org/10.1016/j.cub.2021.10.017
Bunford, N., Andics, A., Kis, A., Miklósi, Á., Gácsi, M., 2017. Canis familiaris As a Model for Non-Invasive Comparative Neuroscience. Trends Neurosci. 40, 438–452. https://doi.org/https://doi.org/10.1016/j.tins.2017.05.003
Cuaya, L. V, Hernández-Pérez, R., Boros, M., Deme, A., Andics, A., 2022. Speech naturalness detection and language representation in the dog brain. Neuroimage 248, 118811. https://doi.org/https://doi.org/10.1016/j.neuroimage.2021.118811
Dror, S., Miklósi, Á., Sommese, A., Temesi, A., Fugazza, C., 2022. Acquisition and long-term memory of object names in a sample of Gifted Word Learner dogs. R. Soc. Open Sci. 8, 210976. https://doi.org/10.1098/rsos.210976
Fugazza, C., Andics, A., Magyari, L., Dror, S., Zempléni, A., Miklósi, Á., 2021. Rapid learning of object names in dogs. Sci. Rep. 11, 2222. https://doi.org/10.1038/s41598-021-81699-2
Fugazza, C., Miklósi, Á., 2020. Depths and limits of spontaneous categorization in a family dog. Sci. Rep. 10, 3082. https://doi.org/10.1038/s41598-020-59965-6
Gábor, A., Gácsi, M., Szabó, D., Miklósi, Á., Kubinyi, E., Andics, A., 2020. Multilevel fMRI adaptation for spoken word processing in the awake dog brain. Sci. Rep. 10, 11968. https://doi.org/10.1038/s41598-020-68821-6
Griebel, U., Oller, D.K., 2012. Vocabulary Learning in a Yorkshire Terrier: Slow Mapping of Spoken Words. PLoS One 7, e30182.
Hare, B., Tomasello, M., 2005. Human-like social skills in dogs? Trends Cogn. Sci. 9, 439–444. https://doi.org/10.1016/j.tics.2005.07.003

Higaki, F., Faragó, T., Pogány, Á., Miklósi, Á., Fugazza, C., 2025. Sound quality impacts dogs' ability to recognize and respond to playback words. Sci. Rep. 15, 14175. https://doi.org/10.1038/s41598-025-96824-8

Kaminski, J., Call, J., Fischer, J., 2004. Word learning in a domestic dog: evidence for "fast mapping". Science 304, 1682–3. https://doi.org/10.1126/science.1097859

Kőszegi, H., Fugazza, C., Magyari, L., Iotchev, I.B., Miklósi, Á., Andics, A., 2023. Investigating responses to object-labels in the domestic dog (Canis familiaris). Sci. Rep. 13, 3150. https://doi.org/10.1038/s41598-023-30201-1

Magyari, L., Huszár, Z., Turzó, A., Andics, A., 2020. Event-related potentials reveal limited readiness to access phonetic details during word processing in dogs. R. Soc. Open Sci. 7, 200851. https://doi.org/10.1098/rsos.200851

Mallikarjun, A., Shroads, E., Newman, R., 2021. The role of linguistic experience in the development of the consonant bias. Anim. Cogn. 24. https://doi.org/10.1007/s10071-020-01436-6

Mallikarjun, A., Shroads, E., Newman, R.S., 2023. Language preference in the domestic dog (Canis familiaris). Anim. Cogn. 26, 451–463. https://doi.org/10.1007/s10071-022-01683-9

Mallikarjun, A., Shroads, E., Newman, R.S., 2019. The cocktail party effect in the domestic dog (Canis familiaris). Anim. Cogn. 1–10. https://doi.org/10.1007/s10071-019-01255-4

Markman, E.M., Abelev, M., 2004. Word learning in dogs? Trends Cogn. Sci. 8, 479–481. https://doi.org/10.1016/j.tics.2004.09.007

Miklósi, Á., Topál, J., 2013. What does it take to become 'best friends'? Evolutionary changes in canine social competence. Trends Cogn. Sci. 17, 287–294. https://doi.org/https://doi.org/10.1016/j.tics.2013.04.005

Pilley, J.W., Reid, A.K., 2011. Border collie comprehends object names as verbal referents. Behav. Processes 86, 184–95. https://doi.org/10.1016/j.beproc.2010.11.007

Ramos, D., Mills, D.S., 2019. Limitations in the learning of verbal content by dogs during the training of OBJECT and ACTION commands. J. Vet. Behav. Clin. Appl. Res. 31, 92–99. https://doi.org/10.1016/j.jveb.2019.03.011

Venue:

Department of English and American Studies

University of Vienna, Campus, Court 8

Spitalgasse 2, 1090 Vienna