



biolinguistics

volume 14 — special issue — 2020



ISSN 1450-3417

Why Don't Languages Grammaticalize [±poisonous]?

Evelina Leivada^{1,*} & Lluís Barceló-Coblijn²

¹ Department of German and English Studies, Universitat Rovira i Virgili, Spain

² Laboratori d'Investigació en Complexitat i de Lingüística Experimental (LICLE),
Universitat de les Illes Balears, Spain

* Corresponding author: evelina.leivada@urv.cat



EL: <https://orcid.org/0000-0003-3181-1917>

LBC: <https://orcid.org/0000-0002-8765-6314>

1. Introduction

Biolinguistics is the study of language from a biological point of view. As the *Special Issue—Biolinguistic Research in the 21st Century of Biolinguistics* shows, this area of study can be approached from different perspectives: aspects of language evolution (Wacewicz et al. 2020; Mendívil-Giró 2020), ethology and cross-species comparisons (Schalz & Dickins 2020), and neural network analysis (Collins 2020) are some of them. Instead of presenting a brief overview of these works, we will use this editorial article to illustrate a more valuable and often overlooked point, namely, that another perspective to the study of language from a biological point of view is the one that goes through theoretical linguistics. Any study in theoretical linguistics that strives for explanatory adequacy in terms of the innate primitives it assumes falls within the domain of biolinguistics. Although some theoretical linguists explicitly distance themselves from biolinguistics, the question of ‘What is innate in language behavior?’ as well as any discussion that evokes innate characteristics is contributing towards a biological theory of language. This is not an original claim; it is how Eric Lenneberg chose to close the last chapter of his ground-breaking book *Biological Foundations of Language* (Lenneberg 1967).

Some of the questions that Lenneberg addressed regarding innate features and cross-linguistic variation are still looking for answers today. One example comes from the first two live debates in a series of video conferences called “Linguistics Flash-Mobs. Epic Battles in History”, which started in May 2020. In each meeting, two scholars are invited to discuss longstanding theoretical issues



in linguistics. In the first debate, two theoretical linguists, Giuseppe Longobardi and Ian Roberts, addressed the topic of parameters and parametric variation. Roberts posed a critical question that summarizes perfectly (i) our current understanding of the set of features languages grammaticalize and (ii) why these features are selected for grammaticalization instead of others. In his words, “which (sub)set of formal features is underspecified is a question that is very difficult to answer in part because we don’t really know what a set of formal features is” (Roberts 2020). Addressing this matter, Luigi Rizzi argued that “clearly U[niversal] G[rammar] must say something about the set of formal features. A rather elementary empirical observation is that of all the properties that are cognitively salient or perceptually salient only a very small subset is actually grammatical[ized]” (Rizzi 2020).

To sum up, the argument is that languages grammaticalize only a subset of features and the link that Rizzi established is between UG, as the locus of the features that have been selected for grammaticalization, and parameterization, as the process that accounts for the cross-linguistically variable values of these features. Rizzi also gave a useful example to illustrate the difference between grammaticalized and non-grammaticalized features. In his words, “something like physical shape can be grammaticalized, as many systems of nominal categories may refer to shape, but no language, to the best of our knowledge, may refer to color. There is no agreement for yellow objects” (Rizzi 2020). As Ian Roberts observed, one then faces the following question when addressing this topic: Why doesn’t UG encode a color agreement feature?

Before addressing this question—a question which according to Roberts has not received an answer yet—it is useful to bring into the picture the second live debate of the “Linguistics Flash-Mobs. Epic Battles in History” series. This debate featured again two theoretical linguists, Peter Svenonius and Guglielmo Cinque, who discussed an issue that is conceptually related to the topic of the first debate: functional hierarchies. One of the very first points that Cinque raised related to why we find that only some functional elements are encoded grammatically in all languages. He further argued that our cognitive system is quite rich and involves many notions that are not coded grammatically (Cinque 2020). His conclusion, formed on the basis of these observations, echoed Rizzi’s conclusion in the first debate: UG must encode the toolkit of features that languages grammaticalize.

Indeed, studies in crosslinguistic variation attest to the fact that the vast majority of languages grammatically represent certain notions (for example, Tense or Aspect), but not others (for example, whether something is poisonous). The research questions that emerge in this context are the following two:

RQ 1: Does the observation that certain notions are grammaticalized, while others are not, offer an argument in favor of a rich UG that encodes the relevant features?

RQ 2: Why does UG (or language in general, if one does not link the presence of this feature toolkit with UG) grammatically encode specific features, but not others that are equally salient from a cognitive point of view (e.g., color)?

2. A Rich UG?

Undoubtedly, there are restrictions in the pool of features that languages grammaticalize. However, the conclusion that several scholars, including Rizzi and Cinque, reach on the basis of this premise, namely, that UG dictates the set of grammaticalized features, does not follow as a logical conclusion. The observation that certain notions are consistently grammaticalized across languages only entails that languages tend to grammaticalize certain markers, but not others. The reasons for this may have nothing to do with UG. How *Homo sapiens* evolved towards a phenotype that is able to process and encode specific grammatical markers is an empirical question. A rich and highly detailed UG entails a rigid system, and this is less desirable for a developmental process like language development, that follows a nonlinear trajectory and shows abrupt transitions (Ruhland & van Geert 1998; Ninio 2006; Corominas-Murtra et al. 2009; Bassano et al. 2011; Barceló-Coblijn et al. 2012). A system that is flexible enough can reach the mature state more easily than a very detailed, rigid system. From this perspective, it would be surprising if brains show ontogenetic plasticity, but develop rigid cognitive systems. An innate large set of highly detailed features or patterns would also be problematic for language acquisition. When humans have to learn artificial, highly detailed and specific languages, the chosen strategy by the learners tends to be an adaptation and reduction of the features towards under-specification of meanings (Kirby et al. 2008).

Another challenge for the proponents of the position that UG encodes the set of features that languages grammaticalize boils down to the fact that, unlike what RQ 1 may imply, languages do vary as to which features they grammaticalize. For example, if the Tense category is understood as a verb inflection that specifies TIME, spoken languages largely conform with this pattern, but verbs in sign languages often do not inflect for tense. Rather, temporal information is conveyed by time adverbials and/or is inferred from the context (Aarons et al. 1995; Pfau & Steinbach 2006). Clearly, languages vary in terms of the grammatical markers they employ, and they also vary in relation to the strategies they use for expressing the same notion (i.e. grammaticalization or lexicalization). If TIME can be variably expressed across languages through an adverbial, an inflectional marker, or background information, can this notion be removed from the hypothetical UG toolkit that specifies the subset of formal features that are grammaticalized? Put another way, if this notion is expressed in some languages in the *absence* of any dedicated grammatical marker, this feature is potentially grammaticalizable, but not necessarily grammaticalized across all the languages in which it is expressed. It thus seems that RQ 1 must be amended as follows, in order to make reference to notions that are grammaticalizable, but not necessarily grammaticalized: Does the observation that certain notions are grammaticalizable, while others are not, offer an argument in favor of a rich UG that encodes the relevant features? This reformulation of the question puts the matter in its right dimension but brings forward a major problem. It has *not* been shown that features like COLOR or BRIGHTNESS cannot be grammaticalized. What has been observed is that formal grammars do not develop grammatical markers for such notions (Cinque 2013; Adger 2018; Sigurðsson 2020). This suffices to give a

negative answer to RQ 1, because (i) the possibility that COLOR or POISON can be grammatically expressed and the relevant markers successfully acquired has not been discarded and (ii) even if (i) is established, it is not necessary that it is UG that precludes the grammaticalization of certain markers. As the next section will show, several other explanations can be given as to why some features are not grammaticalized, such as adaptation pressures towards input simplification as well as notion-specific semantic restrictions that may favor lexicalization over grammaticalization.

3. Why Not [\pm poisonous]?

The question that emerges next is why languages don't grammaticalize [\pm poisonous] or [\pm bright] if in principle these notions are grammaticalizable. RQ 2 phrases this in a more general way by asking why languages grammatically encode specific features, but not others that may be equally salient from a cognitive point of view. Recall that versions of this question have been voiced as lending support to the idea that if [\pm poisonous] and other notions are not grammaticalized, it must be the case that UG encodes a toolkit of features from which languages variably choose what to grammaticalize (i.e., not all languages have a grammatical marker for evidentiality, although potentially they could in the sense that this is a grammaticalizable, learnable marker).

One answer to RQ 2 is that the semantics of [\pm poisonous] is not a good candidate for an atomic semantic primitive in the first place. It is an endpoint, not a building block.¹ A second answer is that any evolved feature must serve a purpose (like any selected trait for an organism) such as facilitating successful reference in some respect. Having a grammatical marker for poison would not serve any such purpose, so from a biolinguistic point of view it would be superfluous and unlikely to ever develop. Of course, describing whether something is poisonous or not is extremely useful in any community, but languages do not need to express this through a grammatical marker, whereas they often need to grammaticalize past and present tense for obvious reasons of time reference any time an event is described. A third answer has to do with learnability considerations. A grammatical marker that encodes [\pm poisonous] would probably fail to be re-transmitted by learners even in cases of iterated learning in artificial language learning tasks, precisely because it would be useless in most contexts.

A fourth answer relates the existence of formal features with the evolutionary origins of our species. A number of scholars have argued that only concepts that predate the emergence of *Homo sapiens* are candidates for formal features (Emonds 2011; Golston 2018; Panagiotidis 2021). Emonds (2011) presents this hypothesis in the following way.

(1) *Emonds' anti-correlation hypothesis*

Almost all concepts F of human syntax are among those plausibly associated with the cognition of highly developed non-human primates, precisely those who *lack* syntax.

¹ This observation is due to Gillian Ramchand, offered during a Twitter exchange.

This answer seems to suffer from two problems. First, as Emonds (2011) argues, many grammatical features of human language are indeed present in the communication systems of other primates (e.g., ANIMATE). Although this is largely true for animacy, the same cannot be argued for all features that human language can grammaticalize. One example is evidentiality. Defined as the linguistic designation of source of information for a belief, evidentiality has been described as the linguistic coding of epistemology (Chafe & Nichols 1986). Languages differ in the way they grammatically encode their evidential markers across many dimensions. The first and most fundamental difference is that, although all languages have a way to indicate source of evidence for an utterance, only a quarter of them have a grammatical system of specific markers for encoding it (Aikhenvald 2004). These markers can denote source of evidence (i.e., hearsay, inferential, dubitative), modality of source (i.e., auditory, visual), and degree of certainty. According to Emonds' anti-correlation hypothesis (1), the prediction is that such notions *are expressed* in the repertoire of other species. However, the use of the "hearsay" evidential, for example, would entail the transition from a direct predator alarm call (i.e., 'predator X is close') to a "reported speech" predator alarm call (i.e., 'I heard/they say/it is reported that predator X is close'), something that contradicts our current understanding of animal cognition, which has been argued to not provide evidence of hierarchically organized structure learning (Petkov & ten Cate 2020).

The second problem with Emonds' anti-correlation hypothesis is that it predicts that notions that are not grammaticalized in human language must be absent from the cognition of other primates. For example, he argues that is highly doubtful that other species classify the activities of others as respectful, generous, or selfish. Again, results from recent animal studies do not seem to borne out this prediction. For instance, some works suggest that domestic dogs could be able to recognize generous and selfish people and later express this as a systematic preference in choice tests (Carballo et al. 2015). Other works report that chimpanzees could also be able to infer the reputation of humans as selfish or generous agents, perhaps also expressing systematic preferences for generous donors over selfish ones (Subiaul et al. 2008).

Going back to RQ 2, [\pm poisonous] is not a good candidate for a grammatical marker in human language for at least three reasons that are not related to the anti-correlation hypothesis in (1). At the same time, POISONOUS is a good candidate for a lexical item in the repertoire of both other species (e.g., chimpanzees produce different type of grunts when encountering different types of food; Slocombe & Zuberbühler 2006) and humans.

4. Outlook

This work focused on two questions that are often addressed within the realm of theoretical linguistics: 1. Why languages consistently (do not) grammaticalize a subset of cognitive notions, and 2. Whether this observation justifies the conclusion that Universal Grammar must encode a toolkit of grammaticalizable features. The first question received several explanations. Learning considerations, adaptation constraints, and semantic restrictions that may make a notion

a better candidate for lexicalization instead of grammaticalization are some of the reasons that explain why TENSE is a good candidate for the development of a grammatical marker, but COLOR or POISON are not.

The second question was answered negatively: Contrary to popular claims within cartography, it has not been established that it is UG that precludes the grammaticalization of certain markers. Providing the bigger picture, the take-home message is that these are empirical questions that show how theoretical linguistics and biolinguistics are conceptually connected, such that addressing long-standing debates in theoretical linguistics inevitably goes through discussing certain key topics in biolinguistics such as evolution, adaptation, and development.

Author contributions

EL and LBC conceptualized and wrote the paper. Both authors jointly revised the manuscript and approved it for publication.

Acknowledgements

We thank all the authors who participated in the *Special Issue—Bilingualistic Research in the 21st Century of Biolinguistics* and all the reviewers who provided their invaluable service. We also thank Associate Editor Patrick Trettenbrein, who provided help in technical matters, and Editor-in-Chief Kleanthes Grohmann, who assisted in the smooth completion of the special issue. For this editorial, EL acknowledges financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement n° 746652 and from the Spanish Ministry of Science, Innovation and Universities under the Ramón y Cajal grant agreement n° RYC2018-025456-I. The funders had no role in the writing of the study and in the decision to submit the article for publication.

References

- Aarons, Debra, Benjamin J. Bahan, Judy Kegl, & Carol Neidle. 1995. Lexical tense markers in American Sign Language. In Karen Emmorey & Judy Reilly (eds.), *Language, Gesture, and Space*, 225–253. Hillsdale: Lawrence Erlbaum.
- Aikhenvald, Alexandra. 2004. *Evidentiality*. New York: Oxford University Press.
- Barceló-Coblijn, Lluís, Bernat Corominas-Murtra, & Antoni Gomila. 2012. Syntactic trees and small-world networks: Syntactic development as a dynamical process. *Adaptive Behavior* 20, 427–442.
- Bassano, Dominique, Isabelle Maillochon, Katharina Korecky-Kröll, Marijn van Dijk, Sabine Laaha, Wolfgang U. Dressler, & Paul van Geert. 2011. A comparative and dynamic approach to the development of determiner use in three children acquiring different languages. *First Language* 31, 253–279.

- Carballo, Fabricio, Esteban Freidin, Natalia Putrino, Carolina Shimabukuro, Emma Casanave, & Mariana Bentosela. 2015. Dog's discrimination of human selfish and generous attitudes: the role of individual recognition, experience, and experimenters' gender. *PLoS ONE* 10(2): e0116314.
- Chafe, Wallace L. & Johanna Nichols (eds.). 1986. *Evidentiality: The Linguistic Coding of Epistemology*. Norwood, NJ: Ablex.
- Cinque, Guglielmo. 2020. Debate 2: Functional hierarchies (Guglielmo Cinque and Peter Svenonius). *Linguistics Flash-Mobs. Epic Battles in History*, Università degli Studi di Padova, https://www.youtube.com/watch?v=r_6DRMgxI44&t=4453s.
- Cinque, Guglielmo. 2013. Cognition, Universal Grammar, and typological generalizations. *Lingua* 130, 50–65.
- Collins, Joe. 2020. The phonological latching network. *Biolinguistics* 14.SI, 102–129.
- Corominas-Murtra, Bernat, Sergi Valverde, & Ricard V. Solé. 2009. The ontogeny of scale-free syntax networks: Phase transitions in early language acquisition. *Advances in Complex Systems* 12, 371–392.
- Emonds, Joseph E. 2011. From primate to human in two easy steps. In Nicholas Saul & Simon James (eds.), *The Evolution of Literature: Legacies of Darwin in European Cultures*, 43–69. Amsterdam: Rodopi.
- Golston, Chris. 2018. Phi-features in animal cognition. *Biolinguistics* 12, 55–98.
- Kirby, Simon, Hannah Cornish, & Kenny Smith. 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.
- Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: John Wiley & Sons.
- Mendívil-Giró, 2020. In defence of FLB/FLN: A reply to Wacewicz et al. (2020). *Biolinguistics* 14.SI, 145–153.
- Ninio, Anat. 2006. *Language and the Learning curve: A New Theory of Syntactic Development*. Oxford: Oxford University Press.
- Panagiotidis, Phoebos. 2021. Towards a (minimalist) theory of features: Preliminary notes. *Lingbuzz*, <https://ling.auf.net/lingbuzz/005615>.
- Petkov, Christopher I. & Carel ten Cate. 2020. Structured sequence learning: Animal abilities, cognitive operations, and language evolution. *Topics in Cognitive Science* 12(3), 828–842.
- Pfau, Roland & Markus Steinbach. 2006. Modality-independent and modality-specific aspects of grammaticalization in sign languages. *Linguistics in Potsdam* 24, 5–98.
- Rizzi, Luigi. 2020. Debate 1: Three questions on parameters (Ian Roberts and Giuseppe Longobardi). *Linguistics Flash-Mobs. Epic Battles in History*, Università degli Studi di Padova, <https://www.youtube.com/watch?v=uGCHNcI4kGg>.
- Roberts, Ian. 2020. Debate 1: Three questions on parameters (Ian Roberts and Giuseppe Longobardi). *Linguistics Flash-Mobs. Epic Battles in History*, Università degli Studi di Padova, <https://www.youtube.com/watch?v=uGCHNcI4kGg>.
- Ruhland, Rick & Paul van Geert. 1998. Jumping into syntax: Transitions in the

- development of closed class words. *British Journal of Developmental Psychology* 16, 65–95.
- Schalz, Sabrina & Thomas E. Dickins. 2020. Humans discriminate individual zebra finches by their song. *Biolinguistics* 14.SI, 130–144.
- Sigurðsson, Halldór Ármann. 2020. Universality and variation in language: The fundamental issues. *Evolutionary Linguistic Theory* 2(1), 5–29.
- Slocombe, Katie E. & Klaus Zuberbühler. 2006. Food-associated calls in chimpanzees: Responses to food types or food preferences? *Animal Behaviour* 72(5), 989–999.
- Subiaul, Francys, Jennifer Vonk, Sanae Okamoto-Barth, & Jochen Barth. 2008. Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers. *Animal Cognition* 11, 611–623.
- Waciewicz, Slawomir, Przemyslaw Zywczyński, Stefan Hartmann, Michael Pleyer, & Antonio Benítez-Burraco. 2020. *Language* in language evolution research: In defense of a pluralistic view. *Biolinguistics* 14.SI, 59–101.

Language in Language Evolution Research: In Defense of a Pluralistic View

Sławomir Wacewicz^{1,*}, Przemysław Żywiczyński¹, Stefan
Hartmann², Michael Pleyer³, & Antonio Benítez-Burraco⁴

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

² Faculty of Arts and Humanities, University of Düsseldorf, Germany

³ Faculty of Arts and Humanities, University of Koblenz-Landau, Germany

⁴ Faculty of Philology, University of Seville, Spain

* Corresponding author: wacewicz@umk.pl



SW: <https://orcid.org/0000-0003-1488-6220>

PŻ: <https://orcid.org/0000-0002-0267-3845>

SH: <https://orcid.org/0000-0002-1186-7182>

MP: <https://orcid.org/0000-0001-6685-391X>

ABB: <https://orcid.org/0000-0003-4574-5666>

Many controversies in language evolution research derive from the fact that *language* is itself a natural language word, which makes the underlying concept fuzzy and cumbersome, and a common perception is that progress in language evolution research is hindered because researchers do not ‘talk about the same thing’. In this article, we claim that agreement on a single, top-down definition of language is not a *sine qua non* for good and productive research in the field of language evolution. First, we use the example of the notion FLN (‘faculty of language in the narrow sense’) to demonstrate how the specific wording of an important top-down definition of (the faculty of) language can—surprisingly—be inconsequential to actual research practice. We then review four approaches to language evolution that we estimate to be particularly influential in the last decade. We show how their breadth precludes a single common conceptualization of language but instead leads to a family resemblance pattern, which underwrites fruitful communication between these approaches, leading to cross-fertilisation and synergies.

Keywords: language; evolution of language; language faculty; language readiness

Editors: Lluís Barceló-Coblijn, Universitat de les Illes Balears, Spain
Evelina Leivada, Universitat Rovira i Virgili, Spain

Received: 10 May 2020

Accepted: 18 July 2020

Published: 23 March 2021

ISSN 1450–3417

CC BY 4.0 License
© 2020 The authors



1. Introduction

The emergence of linguistic behaviour undoubtedly counts among the few most defining developments in the history of our species. Darwin (1871) considered language to be the greatest invention of humankind, only equalled by fire, and Maynard Smith & Szathmáry (1995) include language on their list of only eight “major evolutionary transitions”, alongside e.g. chromosomes or sexual reproduction. Christiansen & Kirby (2003) call the evolutionary emergence of language “the hardest problem in science”, a label that is as bold as it is useful: Its last part underscores the progression of academic interest in language origins from the spheres of mythology, religion and philosophising to the domain of scientific investigation. Language evolution understood as a field of study (or “the Science of Language Evolution”; Żywicznyński 2018) is an interdisciplinary research field concerned with addressing this problem (see, e.g., the collection of papers in Tallerman & Gibson 2012 for an overview).

However, *language* is itself a natural language word that is fuzzy and polysemous, and as such eludes precise definitions. To a great extent, the same can be said of the diverse conceptualizations and technical uses of *language* promoted by the different branches of linguistics. Many have viewed the elusive nature of the term *language* as an inherent stumbling block to progress, lamenting the fact that researchers do not ‘talk about the same thing’ when discussing the evolution of language (e.g., Wescott 1991, Botha 2000, Hauser et al. 2002; see also Jackendoff 2010, Haspelmath 2016). In particular, many approaches use the term *language* to refer to a socially shared external code and see this as the explanandum of language evolution research, whereas other approaches are interested in language as a cognitive system, and still others as a biological entity, such as a genetically specified faculty or as a component part of the human brain (see Balari & Lorenzo 2016). Attempts to specify this explanatory target with more precise technical terminology have remained unsuccessful, and although the ontological complexity of language admittedly plays a role, to a large extent this is because different theoretical stances presuppose different conceptions of what constitutes language “proper”.

2. Criticisms of the Conceptual Diversity of *Language*

Disagreements on the nature of language have always been present in the modern era of language evolution research, which to many starts in 1996 with the launch of the Evolang conference series, “the major meeting for researchers worldwide in the origins and evolution of language” (<http://evolang.org>). This first meeting resulted in a proceedings volume (Hurford et al. 1998) that was scrutinised by Rudolf Botha, himself one of the pioneers of the Evolang movement. Botha (2000) lists thirteen conceptualisations of *language* that can be found in this single volume: ‘aspect of human behavior’, ‘process’, ‘gigantic meta-task’, ‘special human skill’, ‘activity’, ‘species-specific capacity’, ‘sort of contract signed by members of a community’, ‘hard-wired (individual) competence’, ‘group behaviour of social animals, application of Theory of Mind and social intelligence’, ‘mass phenomenon actualized by different agents interacting with each other’, emergent property

that spontaneously forms itself, and ‘complex system of labels for concepts and conceptual structures’. From this, Botha (2000) concludes that the

[...] profusion of ontologically distinct ways of characterizing language is symptomatic of a foundational flaw in work on language evolution: it indicates the absence of a shared, well-founded linguistic ontology. (2000: 152)

He goes on to suggest that “[t]he linguistic entity or entities whose evolution is at issue should be identified and characterized in a clear and non-arbitrary way”, and that a lack of such a consensus is a fundamental stumbling block to progress:

[I]n the absence of broad agreement about what language is as opposed to other linguistic entities, discussions [of language evolution] are bound to deal with questions of language evolution in ways that are inconclusive and internally disconnected.¹

(Botha 2000: 149, 152–153)

A similar critique—also catalysed by an Evolang meeting (2002 in Harvard)—was voiced in the consequential *Science* paper by Marc Hauser, Noam Chomsky, and Tecumseh Fitch. The founding premise of these authors was that “[t]he word ‘language’ has highly divergent meanings in different contexts and disciplines” (Hauser et al. 2002: 1570) and is simply too capacious to be productively used in scientific discourse without further specification.³ Hauser et al. (2002) propose a remedy in the form of a more circumscribed and thus supposedly more scientifically productive notion of (the biological faculty of) language, which they term the faculty of language in the narrow sense, FLN, as separate from the faculty of language in the broad sense, FLB. We return to this distinction in section 3; here we underscore that the paper and the distinction essentially resulted from a deep dissatisfaction with the breadth and polysemy of the everyday word *language*, and the resulting diversity in its use in language evolution research. This is particularly evident in later commentaries, where the authors explain that their motivation behind proposing FLN was “to clarify misunderstandings and aid interdisciplinary rapprochement” (Fitch et al. 2005: 179), and

¹ Similar and equally influential criticisms were also voiced before Evolang. For example, “[o]ne of the reasons for the extremely inconclusive outcome of scholarly debates on the origin and evolution of language is that so few glossogonists define language in the same way. Yet because their definitions, in most cases, remain implicit, definitional differences are rarely acknowledged. Before we can distinguish terminological disagreements from substantive disagreements, we must, I think, be as explicit as possible about what each of us means by the word ‘language’” (Wescott 1991: 77).

³ Cf. Bolhuis et al. (2014: 1): “In our view, for the purposes of scientific understanding, language should be understood as a particular computational cognitive system, implemented neurally, that cannot be equated with an excessively expansive notion of ‘language as communication’. [...] In place of a complex rule system or accounts grounded on general notions of ‘culture’ or ‘communication,’ it appears that *human language syntax* can be defined in an extremely simple way that makes conventional evolutionary explanations much simpler.”

[...] to clarify discussion and avoid confusion, once we realized that researchers (including ourselves) had been using the same word, ‘language’, to talk about two different things (FLB and FLN) for many years, and thus had been talking past each other.

(Fitch 2010: 22)

A related but more recent concern was a much-discussed terminological critique by Martin Haspelmath, commenting on the foundation of the *Journal of Language Evolution* (which, next to *Evolang*, was another institutional milestone to language evolution research; see Dediu & de Boer 2016). While the target of Haspelmath’s (2016) commentary is the term *evolution* and its apparent semantic extension from ‘language origins’ to ‘language emergence and language change’, it also targets *language* by extension, since the types of processes that are included under the rubric “evolutionary” predetermine the range of entities they apply to. Haspelmath criticises the resulting lack of clarity and terminological rigour:

As usual, this semantic change of ‘evolution’ has happened because different communities are interested in different concepts, and people (including scientists) are reluctant to coin new terms for new concepts, preferring to adopt old terms from neighbouring communities. Moreover, even linguists tend to be unaware of semantic changes and thus sloppy about terminological use.⁴ (Haspelmath 2016)

In a recent paper, Haspelmath (2020) has coined the cover term *human linguisticity*, understood as a biological capacity which is best studied in a broadly comparative perspective and which imposes some constraints on possible language systems, but with most similarities between languages resulting from convergent cultural evolution. As he explains, this was done to avoid “confusing terminology (‘language faculty’, ‘universal grammar’) [that] has often clouded the substantive issues in the past.”

All in all, these examples show that the terminological fuzziness surrounding the notion of language in language evolution research has been subject to much criticism. In the remainder of this paper, we will defend a position directly opposed to these influential voices, and to a degree contrary to standard intuitions. We will claim that agreement on the ontology of language is not a *sine qua non* for good and productive research in language evolution, and question not

⁴ Following up on Haspelmath’s comment, Mendivil-Giró (2019) argues that the concepts of language evolution and change should be kept apart. He defines language as “a historically modified mental organ” and argues that language change gives rise to language diversity, as new languages can develop out of another language. This process, however, is distinct from the emergence of language from non-language, which, he argues, is beyond the scope of historical linguistics and rather brought about by processes studied in evolutionary biology. However, he also acknowledges that whether or not language evolution and change are seen as a continuum depends on the underlying conceptualisation of language: While he defends the view of language as externalisation patterns of a species-specific Faculty of Language, he concedes that “the conflation of the process of linguistic change and the process of the evolution of FL is natural in those approaches that conceive of languages as social and cultural objects.”

only whether such an agreement may be possible, but also whether it would be necessarily desirable.

As our main point, we question the value of a single top-down notion of language. Such a definition seems not to be possible for language evolution as a field of research both because of the nature of highly interdisciplinary scientific practice in the field and due to the special and multi-faceted ontology of language as an object of study. The only type of overarching definition of language is a bottom-up one, as a family-resemblance notion derived from the patterns of use of the word *language* in everyday language(s) and reflected in the patterns of actual research practice. This leads to definitions of language evolution that could be seen as disappointingly broad and possibly circular, as in “[l]anguage evolution researchers are interested in the processes that led to a qualitative change from a non-linguistic state to a linguistic one” or “[w]e can characterise the study of language evolution as being concerned with the emergence of language out of non-language” (Scott-Phillips & Kirby 2010: 412). However, a ‘usage-based’ understanding of *language* and *language evolution* has the virtue of actually capturing how these terms function in the scientific community of language evolution researchers.

In addition—and again largely as a consequence of the nature of scientific practice—we suggest that conceptual diversity may actually have beneficial consequences. In short, even though there are considerable differences in the understanding of *language* (resulting in differences in the understanding of the explanatory goal of the entire enterprise of language evolution research), this does not necessarily imply incommensurability (see especially Pleyer & Hartmann 2019 for a supporting argument). It also does not necessarily hinder local progress on individual phenomena relevant to the understanding of language evolution, and further, may even be conducive to progress more globally and thus benefit the entire field of research.

3. FLN/FLB

We return to the distinction mentioned in Section 2, between the faculty of language in the narrow versus broad sense (FLN/FLB; Hauser et al. 2002), which for several reasons provides a perfect case in point. Not least among these reasons is the central status of FLN/FLB to language evolution research, and even beyond: it is one of the very few terminological-conceptual exports from the field of language evolution to the study of language and cognition at large, included in important linguistic and interdisciplinary tertiary literature such as *The Stanford Encyclopedia of Philosophy* (Cowie 2008) or *The Encyclopedia of Language and Linguistics* (Tincoff & Hauser 2006). Likewise, the original source of the distinction, the paper by Hauser et al. (2002) already mentioned above is doubtlessly among the most influential works in the field, and probably its most widely cited article (5,716 Google Scholar cites as of 11 April 2020).⁵

⁵ The rest of this section develops an argument originally stated in Wacewicz (2012).

In what follows we will not discuss the content and agenda of Hauser et al. (2002) and the ensuing debates in detail.⁶ Instead, we focus on establishing two points that are central to the rest of our argument:

- (1) The proponents of FLN have defined it twice, in fundamentally discrepant ways: the definition originally formulated in Hauser et al. (2002) and the one later formulated in Fitch et al. (2005) describe two distinct entities (not simply different versions of the same entity).
- (2) This fundamental discrepancy has remained virtually completely unrecognised in the literature, which does not appear to have had major consequences for language evolution research, even in frameworks that explicitly adopt and rely on the FLN/FLB distinction.

In sum, the notion of FLN is a highest-profile case in point, serving to illustrate that even a very fundamental confusion about the explicit, top-down definition of a central notion—the faculty of language—can remain in the background of actual research practice, without readily perceptible detrimental effects.

3.1. *The 2002 vs. 2005 Definitions of FLN*

3.1.1. *The 2002 Definition of FLN*

In their original paper, Hauser et al. (2002) define FLN as the “computational core” of the language faculty: The cognitive subsystem responsible for generating the discrete infinity of linguistic expressions. FLN so defined is a term internal to linguistic theory, and the distinction between FLN and FLB is one based solely on a particular theoretical account of language (and thus of ‘the language faculty’). On this original definition, FLN is one part of the more general faculty of language in the broad sense (FLB), which also includes at least two other major components: the sensorimotor (SM) subsystem and the conceptual-intentional (CI) subsystem, which Hauser et al. (2002) illustrate with examples but do not explain in further detail.

Although later misinterpreted (see below), the definition of FLN as the “computational core” is quite unambiguous, and systematically recurs throughout the paper, for example:

Faculty of language–broad sense (FLB). FLB includes an internal computational system (FLN, below) combined with at least two other organism-internal systems, which we call “sensory-motor” and “conceptual-intentional”. (Hauser et al. 2002: 1569–1570 [italics in the original])

⁶ Readers of *Biolinguistics* are likely familiar with the Hauser–Chomsky–Fitch vs. Pinker and Jackendoff debate (Hauser et al. 2002, Pinker & Jackendoff 2005, Jackendoff & Pinker 2005, Fitch et al. 2005), on which there is a large body of existing commentaries (see, e.g., Parker 2006, Wacewicz 2007, Lewandowska-Tomaszczyk 2008, Balari & Lorenzo 2012, Barceló-Coblijn 2012, Wróbel 2012, Boeckx 2012).

FLB includes sensory-motor, conceptual-intentional, and other possible systems (which we leave open); FLN includes the core grammatical computations that we suggest are limited to recursion.

(Hauser et al. 2002: 1570 [Figure 2, caption])

Faculty of language–narrow sense (FLN). FLN is the abstract linguistic computational system alone, independent of the other systems with which it interacts and interfaces.

(Hauser et al. 2002: 1571 [italics in the original])

It is important to observe that Hauser et al. (2002) discuss the property of ‘uniqueness to humans’ and tag it, clearly and repeatedly, as a *hypothesis* about FLN.

By this *hypothesis*, FLB contains a wide variety of cognitive and perceptual mechanisms shared with other species, but *only those mechanisms underlying FLN—particularly its capacity for discrete infinity—are uniquely human*.

(Hauser et al. 2002: 1573 [emphasis added])

Second, although we have argued that most if not all of FLB is shared with other species, whereas *FLN may be unique to humans*, this represents a tentative, testable *hypothesis* in need of further empirical investigation.

(Hauser et al. 2002: 1576 [emphasis added])

Hypothesis 3: Only FLN is uniquely human.

(Hauser et al. 2002: 1573 [italics in the original])

3.1.2. The 2005 Definition of FLN

As stated above, the 2005 article by Fitch, Hauser, and Chomsky (published as a reply to Pinker & Jackendoff 2005) provides a different definition of the FLN/FLB distinction. Fitch et al. (2005) claim there that FLN is defined as being unique to both humans and language:

[...] given that language as a whole is unique to our species, it seems likely that some subset of the mechanisms of FLB is both unique to humans, and to language itself. We dubbed this subset of mechanisms the faculty of language in the narrow sense (FLN).

(Fitch et al. 2005: 180–181)

We thus made the further, and independent, terminological proposal to denote that subset of FLB that is both specific to language and to humans as FLN. To repeat a central point in our paper: FLN is composed of those components of the overall faculty of language (FLB) that are both unique to humans and unique to or clearly specialized for language.

(Fitch et al. 2005: 182)

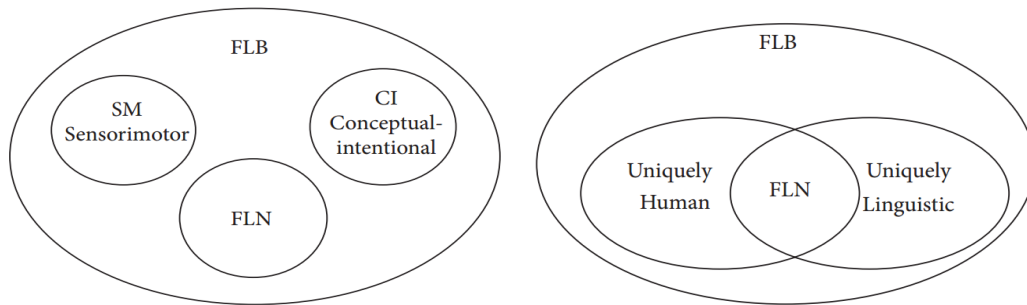


Figure 1: The FLB/FLN distinction in Hauser et al. 2002 (left) and in Fitch et al. 2005 (right).

3.1.3. Summary

In sum, the two papers by Hauser, Chomsky and Fitch reverse the relation between the essential-*cum*-definitional and the accidental-*cum*-hypothetical properties of FLN (Figure 1). The 2002 paper defines FLN as a computational core and hypothesises its human uniqueness; conversely, the 2005 paper defines FLN as uniquely human and hypothesises that it includes a computational core (Table 1). As one example of the consequences, if we accept the 2005 *definition* of FLN as that which is ‘both unique to humans and unique to or clearly specialized for language’, then an unpacked Hypothesis 3 from the 2002 text effectively becomes *only that which is uniquely human (and linguistic) is uniquely human*. This is why—emphatically—the 2002 and 2005 definitions are not just different variants of the same definition, but two different definitions that are discrepant in a strong sense. Unsurprisingly, they produce different answers to consequential questions, such as ‘Can homologous traits be part of FLN?’ Further, they cause a number of literal contradictions, for example:

The contents of FLN are to be empirically determined, and could possibly be empty, if empirical findings showed that none of the mechanisms involved are uniquely human or unique to language, and that only the way they are integrated is specific to human language. The distinction itself is intended as a terminological aid to interdisciplinary discussion and rapprochement, *and obviously does not constitute a testable hypothesis.* (Fitch et al. 2005: 180–181 [emphasis added])

Second, although we have argued that most if not all of FLB is shared with other species, whereas FLN may be unique to humans, this represents a tentative, testable hypothesis in need of further empirical investigation. (Hauser et al. 2002: 1576 [emphasis added])

3.2. Reception and Takeaway

Interestingly, Hauser, Chomsky and Fitch themselves have never addressed the inconsistency and may even remain unaware of it. In Fitch et al. (2005: 181–183), the authors maintain that the later, 2005, definition, and hence the ‘uniqueness to humans’ criterion, was in place in the original paper (which, as demonstrated

	FLN in Hauser et al. (2002)	FLN in Fitch et al. (2005)
Definition	FLN is the core computational mechanism of the FLB.	FLN is the part of FLB that is unique to humans and unique to language.
Hypotheses	<ul style="list-style-type: none"> • FLN is the part of FLB that is unique to humans (i.e. FLN is unique to humans, and no other part of FLB is). • FLN can be equated with recursion. 	Only recursion is unique to humans.

Table 1: Definitions of, and hypotheses about, FLN in Hauser et al. 2002 vs. in Fitch et al. 2005.

above, is incorrect). In later work, they continue to use the term FLN in these two incompatible senses, for example:

FLN—and especially the mechanism of recursion—was defined by Hauser et al. (2002) as a computational process that is responsible for the generative and hierarchical properties of narrow syntax.

(Tincoff & Hauser 2006: 536)

HCF proposed a distinction between the faculty of language in the broad [FLB] and narrow sense [FLN]. FLB is simply those processes of the mind that are both necessary and sufficient to support language. Thus, for example, attention is involved in language processing but is neither unique to language nor unique to humans. FLN includes those processes that are both uniquely human and unique to language... HCF hypothesized that FLN, though potentially an empty set, may only include the computational resources subserving recursion and their interface or mapping to the conceptual-intentional [semantics] and sensory-motor [phonetic] systems. (Hauser et al. 2007: 105)

As for the larger community, the discrepancy seems to have been essentially overlooked in the rather sizable body of commentaries that followed in the wake of the original FLN paper and the debate of Hauser et al. with Pinker and Jackendoff. Most interestingly, the rejoinder by Jackendoff & Pinker (2005) does not clearly expose the discrepancy, but instead proceeds to address the revised definition by Fitch et al. (2005). Other commentators tacitly assume FLN to only have a single definition, sticking to either the 2002 ‘computational core’ definition (e.g. Armstrong & Wilcox 2007, Johansson 2005, Kurcz 2004, Lewandowska-Tomaszczyk 2008) or to the 2005 ‘uniquely human + uniquely linguistic’ definition (e.g., Okanoya 2007, Parker 2006, Számádó & Szathmáry 2006). It is not uncommon to see the 2005 definition incorrectly attributed to the 2002 paper, such as in Samuels (2009: 356): “Hauser et al. (2002) define FLN as those aspects of the language faculty that are unique both to humans and to language”. Occasionally,

commentators equivocate between the two interpretations, conflate them or use them interchangeably without noting their mutual incompatibility (e.g., Kinsella 2009).

In summing up this part of our argument, we wish to underscore that the definitional problem itself, that is the existence of two parallel definitions of FLN, is not unusual in science (since competing definitions of technical terms are commonplace) and is tangential to our present interests. Again, the point we make here is not that the definitional discrepancy exists; rather, our point is how the discrepancy has continued to go *essentially unnoticed*. However surprising—and contrary to an earlier analysis by one of us (Wacewicz 2012)—this provides a striking demonstration that the specific wording of the top-down definitions of language was inconsequential to the research practice of the field.

In the next section, we will focus on a number of approaches that have led to significant progress in the field regardless of the fact that they do not directly map onto the different definitions of FLN/FLB. Instead, they represent ‘multi-component’ approaches to language evolution (cf. Fitch 2017, Benítez-Burraco & Progovac 2020) that outline important aspects of ‘language’ and ‘language evolution’ and eschew the kind of restrictive definitions outlined in Section 3. As such, these approaches are further evidence of our view that the definitional discrepancies discussed above did not hinder progress in the field. Instead, these approaches show the importance not of top-down definitions of language, but of focusing on particular aspects of and hypotheses about language and investigating their relation to other factors relevant to language and its evolution (see also Roberts et al. 2020).

4. *Language as a Family Resemblance Category in Language Evolution Research*

In this section, we flesh out our argument with a brief survey of presently influential lines of language evolution research. Our main goal here is to illustrate the current breadth of the field and show how this breadth brings with it conceptual diversity as an inevitable consequence. We propose a categorisation into four general approaches that constitutively differ in how they conceptualise language, where ‘constitutive’ means such differences that preclude a neat grouping under a single common definition. Nevertheless, we wish to show that these conceptualisations are not entirely disjunct but are in fact characterised by patterns of overlapping similarities—in other words, these uses of *language* form a family resemblance category. As an additional point, we also mention some benefits of this conceptual diversity, i.e. ways in which these four approaches have been mutually valuable and invigorating. We will first discuss conceptualisations of language as a multimodal phenomenon (Section 4.1), before turning to approaches that treat language as a complex adaptive system (Section 4.2). We will then discuss approaches that see language as a form of social interaction (Section 4.3), and finally we will explore approaches that look at language from the perspective of the language-ready brain (Section 4.4).

Since, as we emphasise in Section 5, language evolution is a fast-changing field, we focus on approaches that we consider as particularly prolific and

impactful in the last decade. As such, these approaches also represent trends likely to further gain in importance in the near future (see also Nölle et al. 2020). However, such a selection can never aspire to being fully objective, and in particular our temporal perspective means leaving out foundational work by highly prominent but early language evolution scholars, such as Bickerton (1990), Dunbar (1996), Deacon (1997) or Jackendoff (2002). Such works were central to the inception of language evolution as a science, and are still used as reference points for the discussion of specific topics in the modern day science of language evolution—for instance, Bickerton and Jackendoff often feature in debates about the nature of protolanguage (e.g., Fitch 2010), and Dunbar’s views are referenced with regard to the problem of the social preconditions of language emergence (e.g., Dor et al. 2014, Zlatev 2014).

4.1. *Language as a Multimodal Phenomenon*

One approach that has become ever more popular in the last decade sees language as a multimodal phenomenon. As Vigliocco et al. (2014) forcefully argue, “speech signals are invariably accompanied by visual information on the face and in manual gesture” (Vigliocco et al. 2014: 1). Such a view differs considerably from more traditional conceptualisations, on which non-verbal behaviour (e.g., as defined and taxonomised in the seminal paper by Ekman & Friesen 1969), and primarily gesture, supports but is definitionally separate from linguistic communication. This definitional framework has very profound consequences for language evolution: Since gesture and occasionally other forms of non-verbal communication are inseparable from (spoken) language, it follows that even though the nonverbal component and the verbal component may be analytically distinct, their evolutionary origins constitute an indivisible explanatory target. In other words, at least for the purposes of explaining its evolutionary origins, gesture must be considered as an integral part of language. In the remainder of this section, we illustrate the view of language as a multimodal phenomenon by discussing three exemplary frameworks, focusing on the question of how language is conceptualised in each of these approaches: Adam Kendon’s idea of languaging, David McNeill’s growth point, and Jordan Zlatev’s mimesis hierarchy.

4.1.1. *Adam Kendon: Languaging*

Adam Kendon rejects the traditional idea that a language forms an abstract system of rules (as proposed by e.g. de Saussure 1916), and opposes the view that language has systemic properties. Instead, Kendon (e.g. 1990, 2004) takes a dynamic, usage-oriented view, whose roots can be traced back to Humboldt’s *energeia* and more recently to Goffman’s interactionism. Kendon (e.g. 2014a, 2017) sometimes uses the term *languaging*, to underline the dynamic character of language, or *gesture-speech ensemble* (Kendon 2004: 108), to underline its multimodal character. He argues that language involves “the mobilization of several different semiotic systems in different modalities and deployed in an orchestrated relationship with one another” (Kendon 2014a). The most linguistic element of this orchestration is speech, which has linear structure and is organised by the morpho-

syntactic component (Kendon 2014a). However, as Kendon insists, “the ‘natural’ state of spoken language” (Kendon 2014b: 76) is the context of physical “co-presence”, in which the transmission of meaning, both propositional and non-propositional, depends on speech being coordinated with “extra-oral bodily action”—hand and arm movements, postural shifts, eye contact or facial expressions (Kendon 2004, 2011). The traditional focus on the systemic properties of language results from abstracting it from this “natural” state, and language so construed is of secondary importance both in ontogenetic and evolutionary terms (Kendon 2014b: 72).

The basic unit of language (or alternatively, ‘*linguaging*’) is the utterance, which is the coming-together of speech and extra-oral visible action to translate ideas into “observable behavior, which may be read by others as reportive of those ideas” (Kendon 1980: 208; see also Kendon 2004). The meaning, including propositional meaning, of an utterance results from an interplay of speech and extra-oral bodily visual actions. Importantly, Kendon sees extra-oral visible bodily action as so closely coordinated with speaking that it has to be understood as an integral component of language. The idea of language as the interaction and co-expression of speech and body movement is of crucial importance to Kendon’s view on language origins. He subscribes to a uniformitarian hypothesis, according to which “the early steps of language evolution also consisted of multi-modal signals, instead of being predominantly hand-based or vocalization based” (Kendon 2014b: 69). Hence, he looks for such an evolutionary context that could explain an early integration between vocal-auditory and visual-bodily semiotic resources and argues for a praxic origin of language—in other words, he argues that language is rooted in concrete actions. Accordingly, there was one, albeit complex, executive system for oral-laryngeal and manual action, which served such purposes as mastication and food-handling (cf. MacNeilage 2008). Later, this system was rededicated “in the service of communicative action” (Kendon 2014b: 72): Articulated vocalisation developed early in the hominin line to manage and maintain complex social relations; gesture, understood broadly as deliberate and expressive-communicative movement, also emerged early from the primary practical, manipulatory function of the hand and forelimb. On Kendon’s view, the common origin of speech and gesture (as defined above), as well as the same evolutionary trajectory (from the praxic to the communicative function), explains what he refers to as “the ‘natural’ state of language” (see above):

[...] gestures that are so often a part of speaking are neither supplements nor add-ons. They are integral to speaking. They are so because they are derived from practical manipulatory actions from which speaking itself is also derived. Looked at in this way, we can better understand why it is that visible bodily action is mobilized when speakers speak and why, more generally, when language is used in co-present interaction it always involves *poly-modal* forms of action.

(Kendon 2014b: 75)

To support his scenario, Kendon extensively appeals to research on the integration of speech and gesture in the communication of modern humans, but also to developmental and neurocognitive evidence, such as the coordination of syllabic babbling with hand movements (Kendon 2014a, cf. Ejiri & Masataka 2001) or the role of Broca's area in controlling hand movements as well as movements of the expressive muscles of the face (Kendon 2014b: 69, cf. Willems et al. 2007, Aboitiz 2012). Kendon argues against the view that asserts continuity between ape gestural communication and modern human gestures, which has been emphasised by gestural accounts of language origin (e.g., Hewes 1977, Arbib 2012, Corballis 2013).

4.1.2. David McNeill: *Growth Points*

In McNeill's model, speech and gesture synergistically express the same overall meanings while remaining semiotically distinct and responsible for the transmission of different aspects of the message: speech for propositional content and gestures for imagistic content. According to McNeill, the *stroke* (i.e. the most pronounced phase) of a gesture accompanies the semantically most prominent element of the utterance. In this way, the Growth Point, the basic unit of thinking, becomes externalised. Here, McNeill departs from Kendon's account of modern human communication, which does not posit a categorical division of labour between speech and gesture, but rather argues for their functional interplay, for example, gesture can transmit propositional aspects of meaning (see above), while speech includes vocal means of expressing emotional-imagistic content, as in the case of paralinguistic features (e.g., emotional prosody) or iconic vocal phenomena, as in ideophones, phonesthemes, reduplication or word lengthening (Kendon 2008). They also disagree about the definition of gesture. McNeill (1992, 2012) would further limit (prototypical) gestures (i.e. co-speech gestures) to spontaneous and idiosyncratic hand and arm movements that are functionally integrated with speech. As we have seen, Kendon's understanding of gesture extends beyond the category of co-speech gestures and embraces any deliberately communicative bodily movement (hence, the use of the term 'kinesic'), including postural shifts, eye contact or facial expressions (Kendon 2004, 2011).

The idea of a tight integration between spoken messages and co-speech gesture is also central to McNeill's theory of language evolution, the critical moment of which is the integration of gestural and vocal communication, both at the level of cognition and expression (McNeill 2012). The claim is that language originated from the coming together of vocalisation and gesture to form a propositional-imagistic dialectic. Like Kendon, McNeill submits a uniformitarian explanation as the rationale of his hypothesis: language in its beginnings was qualitatively similar to what it is now; but it should be remembered that he proposes a more limited view of what language is than Kendon. The critical element in the formation of the propositional-imagistic dialectic was the 'twisting' of mirror neurons, whereby they began "to respond to one's own gestures, as if they were from someone else" (McNeill 2012: 65). To support this idea, McNeill paraphrases Mead (1974): "[A] gesture is a meaningful symbol to the extent that it arouses in the one making it the same response it arouses in someone witnessing it" (2012: 180; cf. Arbib's

parity requirement in Section 4.4.1 below). As this gestural system was co-orchestrated with vocalisation, the Growth Point emerged.

It should be noted that McNeill does not provide any evolutionarily grounded pressures that could have been responsible for these changes. In fact, he ventures two rather different accounts of how speech started, deriving it either from ingestion, which assumed vocal properties and was subsequently orchestrated with gesture (2012: 180–181), or from the type of communication that is found in extant non-human apes, such as “chimpanzee gestures with vocalization” (2012: 195). Although McNeill refers to the ‘twisting’ of mirror neurons and the voice-gesture integration as adaptations, he actually describes them as saltational leaps, not unlike Chomsky’s idea of a lucky mutation giving rise to the operation of Merge, which first endowed humans with a language of thought and then with the communicative use of it (Berwick & Chomsky 2016).

4.1.3. Jordan Zlatev: *The Mimesis Hierarchy*

A different account of language and language evolution is put forward by Zlatev (2008, among others). Zlatev objects to the very term ‘multimodality’ as used by Kendon and McNeill (but also many other researchers, see e.g., Wacewicz & Żywiczyński 2017); for him, language and gesture are two distinct semiotic systems, i.e. systems of signs and relations between them (Zlatev et al. 2020), which are characterised by different design features. For instance, linguistic signs are mainly conventional, gestural signs mainly iconic; the syntagmatic relations between linguistic signs are compositional, between gestural signs, they are linear; language uses double articulation, gesture does not. Next, language can utilise different modalities: Vocal in the case of speech, material in the case of writing, bodily in the case of signed languages, and so on (Żywiczyński & Zlatev, in press). Hence, face-to-face communication is typically both polysemiotic, that is it makes use of different semiotic systems (most importantly, language and gesture), and multimodal, that is it makes use of different communication channels (most importantly, vocal for speech and bodily for gesture; Zlatev 2019).

Zlatev’s key theoretical concept is mimesis, adapted from Donald (1991, 2001). His most recent definition of bodily mimesis is the following:

[...] [A]n act of cognition or communication is an act of bodily mimesis if: (1) it involves a *cross-modal mapping* between exteroception (e.g. vision) and proprioception (e.g. kinesthesia); (2) it is *under conscious control* and is *perceived by the subject to be similar* to some other action, object or event, (3) the subject *intends* the act to stand for some action, object or event *for an addressee*, and for the addressee to recognize this intention; (4) it is *not fully conventional and normative*, and (5) it does *not divide (semi)compositionally* into meaningful sub-acts that *systematically relate* to other similar acts, as in grammar. (Zlatev 2014: 206)

On this basis, Zlatev proposes an evolutionary and developmental model known as the *mimesis hierarchy* (Zlatev, 2008). The rudimentary form of *proto-mimesis*, based on requirement (1), is found in activities like emotional and

attentional contagion, and is common for all primates. The more advanced form of *dyadic mimesis* (based on 1 and 2) involves volition and imitation, but not true representation or sign-function; it is common for all great apes. Only at the next level (based on 1, 2 and 3), referred to as *triadic mimesis*, do mimetic acts gain a clear sign-function, as well as Gricean communicative intentions (i.e. that the addressee should understand that a communicative act is being performed for their benefit). Further, point (4) distinguishes mimesis from a conventionalised proto-language and point (5) from language proper.

This provides a useful conceptual apparatus, but does not answer key questions such as what drove the evolutionary process, as well as more specific aspects of how the transition from triadic mimesis (i.e. pantomime) to protolanguage and language took place, including the shift from a dominance of gesture to a dominance of vocalisation. Zlatev (2016) addresses these gaps, but in a somewhat schematic manner. With respect to evolutionary pressures, Zlatev appeals to an increase of pro-sociality in hominins (cf. Tomasello 2008), which might in turn have been ecologically driven by the reproductive strategy of cooperative breeding, where the biological parents receive help in rearing their young from the wider group (Hrdy 2009). Concerning the gradual transition to vocalisation, this is sought in the nature of pantomime itself: a hybrid system that is polysemiotic (i.e. combines various sign and signal systems) and multimodal (i.e. involves different sensory channels). The dominant semiotic system in pantomime is claimed to have been *robustly iconic gesture* (cf. the notion of primary iconicity; Sonesson 1997). The transition towards language entailed a gradual loss in iconicity along various parameters (see Zlatev et al. 2020 for details). Zlatev (2016) attempts to motivate the gradual transition from gesture to vocalisation when the need for less iconicity and more ‘arbitrariness’ arose.

But while language (realised as speech, writing or signing) may be the dominant system in modern human communication when it comes to expressing propositions and narratives, it is rarely used alone, but alongside other semiotic systems such as gesture and depiction (e.g. Green 2014): Polysemiotic communication. An advantage of the mimesis/pantomime approach is that it can help explain this, as pantomime consisted of gesture, vocalisations as well as ‘proto-drawing’, when gestures left marks on surfaces such as sand (Zlatev 2019, Zlatev et al. 2020).

4.1.4. *Language as a Multimodal Phenomenon: Taking Stock*

The defining feature of the views on language discussed above is its multimodal character. However, the term ‘multimodality’, especially as used by Kendon and McNeill, conflates multimodality itself, i.e. the use of different sensory modalities, with polysemioticity, that is the use of different sign systems, most importantly speech and gesture. Beyond the general consensus that language is multimodal, there are differences in the way these approaches account for language and its separability from other semiotic systems. For McNeill and Kendon (cf. the latter’s idea of languaging), language and gesture are two manifestations of the same system—importantly, this system is at its core both communicative and cognitive. Zlatev enumerates criterial attributes of language, in contradistinction to the

critical attributes of gesture, but emphasises that human-specific communication is nevertheless inherently multimodal and polysemiotic; hence, language should be seen as cognitively distinct from the semiotic system of gesture and other semiotic resources but is inseparable from them in actual communicative behaviour. In this regard, his position is similar to that of Levinson (see Section 4.3 below).

All of these authors underline that language is species-specific, but they also posit its continuity with ape cognition and communication, albeit with various degrees of emphasis. They also agree about the watershed in the evolution of language, which was of semiotic nature and consisted in the emergence of iconic gestures, although again, they differ in their account of how abrupt the semiotic breakthrough was. Related to that point is the division of labour between biological and cultural evolution. In multimodal approaches, the bulk of biological preadaptations for language, mainly related to the organisation of the neural infrastructure, happened prior to the semiotic breakthrough and facilitated it. The later course of language evolution was almost exclusively the domain of cultural evolution, which led to the emergence of arbitrary symbols and grammar. In accounting for both protolinguistic beginnings of language and its later phases, the multimodal approaches emphasise the importance of cognitive and social factors, for example, the development of complex forms of Theory of Mind, intentionality or cooperation, and treat modern language as integrated in the human socio-cognitive niche. In this regard, they are highly compatible with theories that treat language both as a complex adaptive system and as a form of social interaction.

4.2. *Language as a Complex Adaptive System*

Another influential perspective on language in language evolution is constituted by approaches that view language as complex adaptive systems (CAS) that emerges from social interaction across the timescales of biological evolution, cultural evolution and ontogenetic development (e.g., Steels 2000; Beckner et al. 2009; Kirby 2012). These approaches place different points of emphasis on particular aspects of complex-adaptive processes—some stress their direct relevance to language emergence; others focus on the (socio-)cognitive mechanisms that underlie them. However, we discuss them together as they share the underlying view of language as being multifactorial and dynamic, and whose evolution is channelled by cognitive, interactive-communicational and cultural-historical contexts.

Complex adaptive systems are defined as “processes involving a number of interacting parts which give rise to emergent processes that show the appearance of design.” (Kirby 2012: 590). This idea has gained momentum in both theoretical and empirical approaches in language evolution. On a more theoretical plane, it has been adopted by many practitioners of usage-based approaches, for example, those resonating with Construction Grammar and Cognitive Linguistics (see, e.g., Pleyer & Winters 2014). But it has also been adopted as a framework for computational modelling and behavioural experiments. Both of these domains of empirical research adhere to the concept of Iterated Learning: “[A] particular kind of cultural transmission” whereby “a behaviour arises in one individual through induction on the basis of observations of behaviour in another individual *who acquired that behaviour in the same way*” (Kirby et al. 2014: 108, emphasis in original).

Iterated Learning has informed a number of computational models of language evolution (e.g., Smith et al. 2003, Smith & Wonnacott 2010), and has extensively been used as the paradigm for lab experiments on the emergence of novel communication (e.g., Kirby, Cornish & Smith 2008, Garrod et al. 2010, Tinitis et al. 2017, among many others). Iterated Learning crucially depends on the concept of a *transmission bottleneck*: The number of possible utterances is larger than an agent can observe in their lifetime, which is why language adapts to the agents' learning biases (see e.g., van Trijp 2011). Thus, "language is adapting in such a way as to ensure its own survival through the transmission process" (Kirby 2012: 595). Importantly, it is not only the users of a language but also languages themselves that undergo adaptation (see also Deacon 1997, Christiansen & Chater 2008).

The Iterated Learning model is, in principle, not only applicable to language but also to other cultural artefacts, traditions or communicative codes such as writing systems (e.g., Garrod et al. 2010). However, most interestingly for our present concerns, it operationalizes a certain general concept of language: *Signals from a finite signal space are mapped to meanings from a finite meaning space* (see, e.g., Kirby et al. 2008, Cornish 2010). While this is of course a deliberate simplification for modelling purposes, it bears many similarities with the widespread view of languages as inventories of form–meaning pairs. It can be traced back at least to Saussure's (1916) sign concept and has been adopted explicitly in Construction Grammar and other usage-based perspectives, where constructions, that is pairings of form and meaning/function, are understood as the basic units of linguistic description. Of course, the notion of the centrality of the Saussurean sign and the evolution of its components is shared by a variety of approaches, including Bouchard (2013) and Hurford (1989, 2007, 2012), who himself mentions Construction Grammar as a suitable framework for investigating the evolution of language (see Hurford 2012: 348–362).

A major advantage of the CAS perspective on language is its generality: The evolution of language and of other cultural traits can be investigated in a shared theoretical framework. However, this generality also entails that its explanatory value is limited. This is why, for example, Larsen-Freeman (2017) characterises Complexity Theory as a 'metatheory' that warrants different object theories. Among the more widespread object theories are a number of usage-based approaches to language, some of which adopt a generalized theory of evolution, or adapt ideas from evolutionary biology (see, e.g., Croft 2000, 2011, Ritt 2004). One consequence is that many approaches within this framework do not take an explicit stance on the issue of unimodality vs. multimodality as signs can potentially be constituted through multiple modalities. The Iterated Learning framework programme might not explicitly label language as multimodal, but this approach agrees with multimodal approaches in stressing that other modalities than the vocal-auditory modality are subject to social-interactional, communicative, and learning pressures and played an important role in the evolution of language (Verhoef et al. 2014, Little et al. 2017, Motamedi et al. 2019).

Given the wide range of factors discussed in the emergence of language, language and its development are clearly not conceived of as domain-specific developments but as being part of a broader suite of cognitive and interactional processes, although this distinction is generally becoming increasingly blurred in a

number of approaches (cf., e.g., Pleyer & Hartmann 2019). From a Complex Adaptive Systems perspective, language involves a multitude of different cognitive and physiological capabilities, including but not limited to embodied cognition (e.g., Bergen 2012), intention reading and pattern finding (e.g., Tomasello 2009), associative memory (e.g., Divjak 2019), and ‘massive storage’ (Hurford 2012: 261). There is therefore no agreement on components that are criterial for language, although social cognitive abilities often take centre stage. This idea is of course not unique to the CAS approach, but it is a stance that has arguably been taken more forcefully in CAS approaches than in most other frameworks.

An important consequence of viewing language from a CAS perspective is that the boundary between cultural and biological factors gets blurred. As Pleyer & Hartmann (2019) have pointed out, this is in line with recent developments in biolinguistics that increasingly adopt an evo-devo perspective (e.g., Benítez-Burraco & Boeckx 2014, Martins et al. 2016, Bowling 2017, see also Section 4.4.2). This is also one important aspect in which the conceptualisation of language as CAS has influenced, and continues to influence, research on language (evolution): Language is investigated on a par with other phenomena that can be seen as results of cumulative evolution—for example, in the framework of cultural evolution theory, which has become increasingly influential in recent years (see, e.g., Richerson & Boyd 2005, Mesoudi 2011). As a consequence, the challenge that language cannot be easily delineated from other phenomena becomes part of a research programme that aims at taking the continuous nature of the phenomena it investigates into account. Even though most approaches that can be seen as belonging to the CAS framework aim at overcoming the strict divide between biology and culture, it seems fair to say that most of them view language, in the first place, as a cultural and communicative phenomenon.

Given that the feedback loop between individual actions and emergent phenomena on a population level is part and parcel of the CAS model, its proponents see language both as an individual and as a supra-individual/social phenomenon, even though different approaches may emphasise one of these two aspects more than the other. In this regard, it is also quite instructive to take a look at the brief history of Construction Grammar, which originally took “a synchronic and mentalist perspective” (Hilpert 2013: 1) by trying to describe the linguistic knowledge of individuals. However, Construction Grammar increasingly widened its scope to models of language variation and change, which makes a population-level perspective necessary. As the emergence of structure is a dynamic, cultural process, there are in principle no categorical distinctions between language and non-language. So, although CAS approaches assume continuity between language and other forms of communication, language is usually seen as species-specific in the sense that the different components that make up the ‘mosaic’ of language may also be found in other animals, but they are only fully in place in modern humans (Elman 1999).

CAS approaches more generally have arguably had a substantial impact on the research landscape in the field of language evolution and the CAS approach in general has been widely adopted (from Steels’ pioneering 2000 article, to Beckner et al.’s widely-cited 2009 paper, to Kirby’s 2012 handbook article), probably partly due to its compatibility with a very broad spectrum of approaches: While

its foundational assumptions have been widely shared across various approaches in language evolution research for a long time, the CAS framework offers a convenient terminological toolkit for making these assumptions explicit, and it also invites researchers to broaden the scope of their research by putting the phenomena they investigate into perspective, which, as we have seen, also has consequences for the question of how a notion like *language* is conceptualized.

A second reason why CAS approaches are gaining popularity is not theoretical by nature, but empirical, as this framework can accommodate more easily than others increasing evidence of complex feedback effects among a variety of factors—physical, cognitive, behavioural, environmental—involved in language evolution, acquisition, and use. Among others, we can mention the constraints imposed on language form (sounds, morphology) by physical (temperature, humidity) and cultural factors (population size, topology of social networks; Lupyan & Dale 2010); or the differential impact of language features on cognitive abilities (such as working memory in Amici et al. 2019). Ultimately, CAS approaches fit better than others with views of human evolution that see the emergence of modern cognition and culture as the result of a complex feedback loop between our biological endowment and our cultural practices, instead of as the outcome of a linear evolutionary process, with modern cognition appearing first and modern culture/behaviour happening later. In the case of language evolution, it is now viewed as the outcome of a feedback loop (seemingly ongoing) between our biological language-readiness, faculty of language, or linguisticity, and our language-supported cultural practices. This ultimately entails that the boundaries between language evolution—understood as the processes that give rise to fully-fledged human language—and diachronic language change become blurred as the biological processes that lead to the emergence (and further development) of language cannot be neatly separated from the cultural ones. Smith (2018) mentions two areas where the assumption of a close interaction between culture and biology seems plausible: On the one hand, he summarizes de Boer's (2000) modelling work on phonological niche construction in the evolution of vowel systems, in which a selection pressure for individuals with more fine-grained articulatory or perceptual capabilities interacts with cultural-evolutionary pressures that continuously push the vowel system to the limits of the available articulatory or perceptual space. On the other hand, Smith (2018) argues that “process of gene–culture co-evolution might also act to constrain cultural evolution, by imposing biological constraints on the kinds of systems which can be learned”, which may be particularly relevant for the evolution of syntax.

A recent promising spin-off of CAS approaches is the self-domestication hypothesis of language evolution, also adopted by Kirby (Thomas & Kirby 2018). In a nutshell, the existence in humans of features of domesticated mammals compared to wild extant primates is claimed to account for both the emergence of a modern language-ready brain, mostly via a biological mechanism, and of modern languages, endowed with all the features that are familiar to linguists, mostly via a cultural mechanism. The cornerstone of the hypothesis is the reduced reactive aggression and the increased social tolerance brought about by self-domestication, which favoured language teaching and learning, and ultimately, the complexification of linguistic forms. However, brain and cognitive changes are also

expected, either directly, as a consequence of domestication processes, or indirectly, via a feedback effect triggered by the new social environment and the new language forms and uses (see Benítez-Burraco 2020, Benítez-Burraco & Progovac 2020 for details).

In sum, the CAS approach thus provides a convenient metatheory that is compatible with a variety of different approaches to language and its evolution, including the view of language as a form of social interaction, which will be discussed in the next section. As it is applicable both to cultural and to biological systems, it is also very much in line with recent approaches that reject a categorical divide between culture and biology.

4.3. *Language as a Form of Social Interaction*

In explaining the emergence of human language, many scholars stress the social embedding of language, and its resultant importance for language evolution. The centrality of social cognition has a long history in the research on the evolution of language and mind (e.g., Byrne & Whiten 1988, Cheney & Seyfarth 2007, Dor et al. 2014, Dunbar 1993, Tomasello et al. 1993). Here, we single out two conceptions that have proved seminal within this perspective on language and its evolution, namely the shared intentionality framework by Michael Tomasello and colleagues (e.g., Tomasello et al. 2005; Tomasello 2008), and Stephen Levinson's (2006) 'human interaction engine' hypothesis. Although Tomasello and Levinson do not expressly formulate a theory of language—the former concentrates on the prerequisites of language; the other, on the problem of human-specific interaction—they definitely subscribe to the vision of language as a form of social cognition and social action, providing the empirical evidence and theoretical scaffolding for this vision.

4.3.1. *Tomasello's Shared Intentionality and Levinson's Interaction Engine*

Tomasello's shared intentionality framework (e.g., Tomasello et al. 2005) certainly ranks among the most important approaches in language evolution, even though one might wonder whether the framework actually seeks to explain the emergence of language or rather the evolution of the cognitive prerequisites for language. The answer to this question depends, again, on how exactly we define *language*. It therefore makes sense to first take a look at Tomasello's conceptualisation of language. While he does not provide a formal definition of *language*, he makes it clear that he sees language as a form of social action (Tomasello 2008: 342–345): "What is language if not a set of coordination devices for directing the attention of others?" (Tomasello et al. 2005: 690). On this view, language can be described as a way not only to coordinate attention, but also to construe objects and events from a particular perspective. But although language itself might aid in the cognitive development of perspective-taking (e.g., Lohmann & Tomasello 2003), the cognitive and interactional machinery it is built on must have emerged prior to language. As Tomasello puts it:

If we want to understand human communication, [...] we cannot begin with language. Rather, we must begin with unconventionalized, uncoded communication, and other forms of mental attunement, as foundational. (Tomasello 2008: 59)

Tomasello (2008) sees language as a human-specific form of coded communication that uses conventionalized ('codified') signs, in contrast to uncoded communication making use of spontaneous, ad-hoc signs (cf. Arbib's pantomime). But importantly, there is no clear dividing line between these two modes of communication, which is why they cannot be discussed in isolation. In this regard, Tomasello's conception of language is similar to the ones espoused by the proponents of the multimodal view. The main difference between them lies in that while Tomasello emphasises the social grounding of language, Kendon, McNeill, and Zlatev tend to focus on the cognitive-interactive dynamics of linguistic communication and assess language and its evolution from this vantage point.

Tomasello and his collaborators see the cognitive infrastructure supporting shared intentionality as being central to human cultural cognition. Shared intentionality can be described as the motivation and ability to engage with others in collaborative activities with joint goals, plans and intentions and to share attention, experiences and other psychological states with others (cf. Tomasello et al. 2005, Tomasello & Carpenter 2007, Tomasello 2008). The shared intentionality infrastructure is hypothesised to be the foundation of uniquely human sociality and cumulative culture. This 'we-perspective' (Tuomela 2007) is what enables humans to have not only an understanding of shared goals in the way that other animals do not but represents the foundation of distinctive human cultural artefacts. These include institutional realities such as money, as well as conventions and norms more generally. They are seen as crucial for the emergence and acquisition of language, understood as a cultural artefact that is both conventional and subject to cumulative cultural evolution as evidenced in language change (Tomasello 1999, 2008, 2019; cf. the view of language as a cumulative technology, Dor 2015). In this way, the shared intentionality infrastructure is seen as the *sine qua non* for the evolution, acquisition, and use of language. Tomasello et al. (2005) acknowledge that language "must play a central role in all discussions of the evolution of human cognition." However, as they stress,

[...] saying that only humans have language is like saying that only humans build skyscrapers, when the fact is that only humans (among primates) build freestanding shelters at all. Language is not basic; it is derived. (Tomasello et al. 2005: 690)

Accordingly, language is seen as part and parcel of other human-specific social and cognitive skills and motivations that are already evident in the nonverbal communicative and cooperative behaviour of prelinguistic infants: the capacities that lead infants to communicate informatively and declaratively.

With regard to the evolutionary trajectory of language emergence, Tomasello agrees with proponents of multimodal approaches (see Section 4.1, but also Section 4.3.2 below) that gesture played a pivotal role in language evolution.

Specifically, he argues that the socio-cognitive infrastructure of shared intentionality was directly responsible for two original forms of human communication – pointing and pantomiming. Tomasello (2008) proposes declarative pointing, and especially informative-declarative pointing (i.e. pointing performed with the intention of providing the recipient with new information) to be the first step distinguishing human ancestors from the generalized baseline of the last common ancestor (LCA) that humans shared with chimpanzees. Pantomiming, the other rudiment of the original human communication system, is understood as an iconic and action-based representation of an event (Tomasello 2008). In this respect, Tomasello’s account ties in with those by Zlatev and Arbib (cf. Sections 4.1. and 4.4) but differs especially from the former in that Tomasello does not assume that pantomime must have been multimodal.

A similar idea, partly drawing on Tomasello’s account, can be found in Stephen Levinson’s ‘interaction engine’ hypothesis (e.g., Levinson 2006), according to which what evolved in our ancestors was a socio-cognitive adaptation allowing “joint attention, common ground, collaboration and the reasoning about communicative intent” (Levinson & Holler 2014: 369). Levinson argues for a model of language evolution in which elements characteristic of modern language incrementally grew upon each other. Levinson designates a package of these elements as the Human Interaction Engine (Levinson 2006). Its key properties include intention-attribution, i.e. responses are to intentions, not to behaviours; cooperation, understood along the Gricean lines; turn-taking; predetermined sequential structures, such as adjacency pairs (cf. Pomerantz 1984); and multimodality, whereby

[...] face-to-face interaction is characterized by multimodal signal streams—visual, auditory, and haptic at the receiving end, and kinesic, vocal, and motor at the producing end. (Levinson 2006: 46)

The Interaction Engine represents “a human interactional specialization”, which is universal across cultures, evolved prior to language and played a key role in its emergence (Levinson 2006: 42).

In contrast to the iterated learning framework and the research programme of Tomasello and colleagues, but in line with the multimodal approaches discussed in Section 4.1, Levinson explicitly stresses that human language is multimodal. For Levinson, language is one part of “human multi-modal communication” (Levinson & Holler 2014), which constitutes one integrated multimodal communication system, not in the sense of a special module or a ‘language organ’, but rather an assemblage of interrelated socio-cognitive abilities and communicative behaviours (Levinson 2006: 54 and *passim*). The assorted nature of human communication is reflected in its evolutionary history—“human communication is evolutionarily stratified, composed of layers of abilities of different types and different antiquity” (Levinson & Holler 2014). The bottom layer is constituted by what he refers to as the ‘ethological elements’, such as mutual gaze or turn-taking, including vocal turn-taking, which can be found in many primate clades (Levinson & Holler 2014). Levinson argues that cooperation may have been bootstrapped by these ethological elements in the context of novel ecological pressures (e.g., favouring bigger groups), which in turn laid the foundation for the

inferential background of human communication, exemplified by the presumption of cooperation (Grice 1975) and the presumption of relevance (Sperber & Wilson 1986).

Levinson and Holler forcefully argue for the continuity of ape and human gesture (in contrast, e.g., to Kendon, see Section 4.1.). Levinson assumes that ape gesture formed the platform for the development of indexical gestures, and most importantly declarative pointing. Here, Levinson largely accepts Tomasello's (2008) argument of how the evolution of prosociality in the hominin line, encapsulated by "joint attention, common ground, collaboration and the reasoning about communicative intent" (Levinson & Holler 2014), enabled the appearance of gestural indexes. In the next stage, iconic gesture emerged, accompanied by simple referential vocalisations, which gradually assumed the dominant role in the transfer of meaning (Levinson & Holler, 2014).

4.3.2. *Language as Social Interaction: Taking Stock*

Tomasello and Levinson see language through the lens of Theory of Mind and social intelligence. Similar to the perspective of complex adaptive systems (Section 4.2. above), they conceive of language evolution as a multifactorial and emergent process but emphasise the role of socio-cognitive preconditions installed in humans via biological evolution. Tomasello (2003) suggests that the emergence of joint attention and joint action paved the way for the emergence of symbols, while grammaticalization led to the development of complex grammar. He points out that "different aspects of language—for example, symbols and grammar—may have involved different processes at different evolutionary times." (Tomasello 2003: 109). Levinson stresses the canalization of language through multifactorial constraints, with some attractors being "cognitive, some functional (communicational), some cultural-historical in nature" (Evans & Levinson 2009: 446).

Both Tomasello and Levinson see language as a layered 'mosaic' of different features, to use Hurford's (2003) metaphor (cf. Boeckx 2012). On this view, language (evolution) is neither strictly biological nor cultural, but instead characterized by an interplay of both evolutionary and cultural-historical processes (Tomasello et al. 2005, Tomasello 2008). They also converge on two other general points. First, language is first and foremost a communicative device—this tenet leads both Tomasello and Levinson to the appreciation of non-linguistic forms of communication, such as gesture, which forms a bridge between their positions and multimodal hypotheses (see Section 4.1). In particular, they both agree on the key role of iconic gesturing on the early, bootstrapping stages of language emergence.

Second, communication is rooted in social action, which itself is ramified by general-purpose cognitive mechanisms. For instance, Tomasello insists that language is not an object in any meaningful sense of that word, but rather one of the forms of social action. Tomasello does make a distinction between conventional, or 'coded', communication on the one hand and unconventionalised, uncoded communication on the other. However, he also reminds us that much of linguistic communication that makes use of conventionalised codes relies on uncoded aspects of meaning—as a case in point, consider pragmatic phenomena such as deixis and anaphora resolution (see Tomasello 2008: 57–59). Similarly, Levinson

resists the idea of language as abstracted from a rich interactional context constituted of social, cognitive and communicative factors.

4.4. *Language in the Language-Ready Brain*

As a fourth and final group of approaches to language evolution, we discuss those highlighting the neural implementation of the human ability to acquire and use language. We begin with Michael Arbib's notion of the *language-ready brain*, which we see as a promising candidate for integrating many lines of interdisciplinary evidence characteristic of more recent language evolution research. As we will show, the concept of language-readiness has been adopted well beyond the specific framework suggested by Arbib. As an example, we mention one specific theory of how the brain became language-ready, namely Offline Brain Systems proposed by Bouchard (2013). We then move on to a research avenue that adopts this notion but complements this perspective with a relatively greater reliance on genetic evidence.

4.4.1. *Michael Arbib: The Language-Ready Brain*

Michael Arbib's account of language origins holds considerable significance for current research on language evolution, not by being any less controversial than its alternatives, but through its remarkable theoretical completeness and the wide range of interdisciplinary data on which it is based. It began as the Mirror System Hypothesis (MSH; Arbib 2005, 2012, 2016) and now continues as Cognitive Neuroprimatology (CNP; Arbib 2018). In his work, Arbib and his collaborators (see especially Arbib et al. 2018) rely on a broad range of interdisciplinary data, which particularly prominently includes comparative data from extant primates (unlike in many other language evolution accounts, not limited to great apes but extending to macaques and other monkey species) as well as results of research on human visual-bodily communication, including both co-speech gesture and sign languages.

In line with this breadth of the evidential basis of his account, Arbib has an encompassing view of language as an explanatory target in language evolution research, making it compatible with other approaches discussed here: for example, it underscores the immanent multimodality of language (cf. Section 4.1.), the importance of both its formal-structural and social-interactional dimension (cf. Sections 4.1 and 4.3), as well as the division of labour between biological evolution in establishing the cognitive infrastructure for (proto-)language (see Sections 4.3 and 4.4.2 below) and cultural evolution in accomplishing the subsequent transition from protolanguage to full human languages (cf. Sections 4.2 and 4.3). Still, like other approaches, Arbib's MSH-CNP also has its specific focus, which in this case is on the cerebral implementation of language, as is evident in the name of the hypothesis as well as the title of his book-length manifesto, *How the brain got language* (Arbib 2012). Consistent with this focus is Arbib's notion of the 'language-ready brain'. This term is particularly useful in organising the discussion, since it does not inherently prioritise biological or cultural-evolutionary processes

but manages to capture human uniqueness (since “only the human brain is language ready”, Arbib 2012: ix).

MSH-CNP espouses an “Evo-Devo-Socio”-perspective on language evolution:

What evolved (*Evo*) was a language-*ready* brain—not a brain with an innate mechanism encoding a universal grammar (Arbib 2007) but rather one enabling a child to acquire language (*Devo*), but only if raised in a milieu in which language is already present, something which, it is claimed, required tens of millennia of cultural evolution after the emergence of *Homo sapiens* (*Socio*). (Arbib 2018: 7).

It highlights the building blocks that are not themselves (traditionally seen as) linguistic but are necessary for language, most importantly the cognitive infrastructure supporting the sharing of meaning: *Parity* and *imitation*. Parity is the ability to ‘translate’ between production and comprehension, whereby the same signal counts for more or less the same meaning to both the producer and the receiver of this signal (at least on a basic level, which neglects the complexities of pragmatic inference, e.g., Scott-Phillips 2015). Whereas a great majority of accounts of language evolution simply take this fundamental requirement for communication for granted, MSH-CNP offers a detailed account of the neuronal implementation of parity, based on the mirror neuron system (e.g., Arbib 2005, 2012). While imitation is an important component of other language evolution accounts, including Tomasello (2008; see Section 4.3 above), Arbib (2012) stands out by describing a succession of steps in its development: from a mirror-neuron system for grasping and manual praxic actions, through simple imitation, then complex action recognition and complex imitation (CAR&IM), ultimately leading to pantomime—initially of grasping and manual praxic actions, then of actions outside of own repertoire.

Pantomime is a characteristic feature of MSH-CNP. It is pantomime that is responsible for bringing about perhaps the most important qualitative breakthrough, i.e. that of *open-endedness* in communication: “freedom to create novel associations” (2012: 261). Arbib (2012: 219) observes that pantomime has “the ability to create an open-ended set of complex messages exploiting the primates’ open-ended manual dexterity”. This potential to flexibly introduce novel signals for novel messages underwrites two other gains in expressive power otherwise typical of language, that is domain-generality and displacement. Pantomime is domain-general in that it can be used to communicate about many semantic domains (rather than being restricted to, e.g., only predator evasion or food), and it can also express meanings displaced in time and space (not concerning the immediate here and now).

As mentioned above, the scope of the evidential basis and the resulting breadth of the language-ready brain approach gives rise to numerous convergences with the other approaches discussed in this paper. For example, Arbib argues that the neural mechanisms supporting language perception and production were first involved in non-communicative actions such as tool production, much in line with Kendon’s proposal (see Section 4.1 above). According to Arbib’s

hypothesis, the complex imitation of hand movements predates language learning and use and evolved as a form of social sharing of practical skills. One interesting spin-off of Arbib's hypothesis is the claim that language universals are mostly due to cultural factors, with language structure mostly evolving via a cultural mechanism. This resonates with much work in the CAS paradigm (see Section 4.2), and especially by Kirby and others, who come to similar conclusions from a different approach, i.e. computational simulations (e.g., Kirby et al. 2007) later complemented by laboratory experiments (e.g., Kirby et al. 2008).

The concept of language-readiness has also been adopted by other researchers who do not necessarily share Arbib's MSH-CNP account. For instance, Bouchard (2013, 2015), who sees language as a system of signs, including combinatorial signs that underlie syntax (a view that shares many similarities with the Construction Grammar view discussed above), argues that language-readiness is a consequence of the emergence of 'offline brain systems'. These are systems that can be triggered not only by external but also by brain-internal events. These offline brain systems enable a more abstract representational level, which allows for concepts and percepts (or rather: representations thereof) to be linked. He explains the emergence of these systems by an increase in synaptic interactions triggered by a number of interacting developments, for example, the larger brain that entails an increased potential for synaptic interactions, the more globular shape of the brain that affords more cross-modular interactions, and alleles that improve synaptic repair, thus dramatically increasing synaptic interactions (see Bouchard 2015). In addition, he also stresses the importance of bio-cultural coevolution: "the long dependency during infancy feeds more cultural material into these additional brain capacities" (Bouchard 2015). On this view, language, and even language-readiness, can be seen as an exaptation, as a 'side effect', as it were, of other, more general biological (and cultural) developments.

This is only one example that shows that the concept of brain-readiness has become central not only to accounts of language evolution, but also for discussing the nature of language. But Arbib's more specific proposal has remained highly influential as well and has been adopted and further developed in subsequent work. We will now discuss these developments in more detail.

4.4.2. Antonio Benítez-Burraco and Cedric Boeckx: *The Language-Ready Brain Revisited*

Most of Arbib's neurobiological discussion is focused on brain areas related to language in humans and to visual and auditory perception in both humans and primates. At the same time, as far as language in prehistory is concerned, he mostly focuses on stone technologies in different extinct hominins. However, very recently a more detailed view has been emerging of how the hominin brain was genetically modified in the evolutionary history of our species to support processes involved in language. In particular, a series of related papers by Benítez-Burraco and Boeckx (Boeckx & Benítez-Burraco, 2014a, 2014b; Benítez-Burraco & Boeckx, 2015) has outlined a refined approach to Arbib's view of the language-ready brain, as well as Bouchard's approach based on the notion that our more globular brain resulted in enhanced cross-modal thinking. These authors also adopt a multimodal approach, but mostly relate it to Poeppel's claims of the

multifunctionality of the neural mechanisms involved in language processing (see Poeppel & Embick 2005; Poeppel 2012). Likewise, following Bouchard's 'neuro-genetic factors' (e.g., Bouchard 2013: Chap. 4), they are particularly interested in the genetic factors accounting for the changes resulting in our globularity. However, they build on recent paleoneurological and paleogenetic research about changes in the human genome and the human brain (and skull), principally after the split of *Homo sapiens* from the closely related clades of Neanderthals and Denisovans.

The most important reason for this new approach was avoiding what Benítez-Burraco and colleagues saw as an overreliance on speculation in the field of evolutionary linguistics, in particular relying on highly elusive and contentious proxies for language such as 'symbolic behaviour'. Instead, they proposed to focus on the most distinctive and less controversial biological differences distinguishing modern humans from Neanderthals and other extinct hominins. Perhaps the most prominent of such differences is the globular aspect of the human endocranial morphology (Bruner et al. 2003; Neubauer et al. 2010; Gunz et al. 2010, 2012). In their papers, Benítez-Burraco and Boeckx argued that this globularisation of the human brain resulted in a rewiring that improved the connections between subcortical (particularly, the thalamus) and cortical structures, habilitating the neuronal workspace needed for transcending the signature limits of core knowledge systems and ultimately allowing to combine and unify conceptual units that belong to distinct core systems. They further argued that this ability can be equated with the core combinatorial operation in natural language (which may be called Merge by linguists of Chomskyan persuasion), which is at the core of our language-ready brain.

Overall, this is a bridging hypothesis linking considerations of syntax and neuroscience. One reason for this is that this freely combining merging ability is argued to be constrained via its interfacing with other cognitive systems and with the devices involved in speech/gesture production. In addition, compatible with the multimodal approaches to language (see Section 4.1.), this regulation is hypothesised to result from basic neurobiological mechanisms, specifically, from the embedding of high frequency oscillations (e.g., gamma) inside oscillations operating at slower frequencies (e.g., alpha). This embedding ultimately enables the synchronization of distant cortical areas where the diverse core knowledge systems are located, with some subcortical structures, particularly the thalamus, acting, as noted, as a relay centre or switching station connecting the cortical areas. The circuits bidirectionally connecting the thalamus and cortex are at the heart of the language-ready brain and share features of the networks responsible for mind-wandering and inner speech (Gruberger et al. 2011), as well as the top-down attentional regulation network (Miller & Buschman 2013). Likewise, some studies (e.g., Hecht et al. 2013) have related the changes resulting in our language-readiness (particularly, the increase in the ratio between fronto-parietal vs fronto-temporal connectivity from monkeys to apes to humans) to the evolutionary shift from emulation (i.e. a way of copying actions that focuses on the goal rather than the specific movements) to imitation (i.e. a way of copying actions that focuses on the specific movements rather than their ultimate goal).

On Benítez-Burraco and Boeckx's view, the emergence of a language-ready brain does not entail any drastic changes to the generalised primate brain, nor the evolution of entirely new neural devices. Rather, interareal cerebral communication via the synchronizing of spatially distributed oscillations is a generic strategy of the brain, specific neither to humans nor to language. Likewise, most if not all brain areas supporting language are most likely present in other species. Accordingly, the evolution of the language-ready brain essentially involved a change in the dynamic connectivity of the brain resulting from a new anatomical context. Notably, Benítez-Burraco and Boeckx use the available information on ancient genomes to identify some of the genetic changes that may account for the observed differences between hominin species regarding globularisation and cognitive abilities. The candidate genes of interest include *RUNX2* (a master transcription factor during vertebrate development) and several of its effectors; components of two gene networks implicated in vocal learning, clustered around the famous 'language gene' *FOXP2*, and the *ROBO* and *SLITs* effectors; and finally, a set of genes clustered around *AUTS2*, strongly linked to autism (which, incidentally, reinforces the intriguing parallelisms between the autistic mind and the hypothesised Neanderthal mind). Similarly to its neurobiological substrate, also the genetic underpinnings of the language-ready brain are mostly shared with other primates, although some human-specific changes can be also identified, seemingly accounting for the changes in neuronal networks described above.

4.4.3. *The Language-Ready Brain: Taking Stock*

On the language-ready brain view, the presence of compositional, open-ended and domain-general semantics is the most important criterial feature of language. This is thought to rely not on an enhanced, language-specific computational ability, but on an unbounded basic combinatorial ability capable of transcending the limits of core conceptual systems. Other components of language, particularly, some forms of phonology and pragmatics, are assumed to predate this human-specific innovation. Accordingly, speech or interaction through (proto)language with communicative or socializing purposes are thought to have been present in other hominin species, particularly Neanderthals. Overall, this is a genuinely non-modular construal of language. Neural devices involved in language processing are hypothesised to perform basic computations that are recruited for language, but also for other cognitive processes. Accordingly, the impairment of any of these neural components of language—either developmentally, resulting from gene mutations, or in the mature state, resulting from brain damage—is expected to give rise to mixed symptoms and diverse pathological, comorbid conditions. As a consequence, too, language is construed as domain-specific only at the term of growth, with the 'language module' resulting from the interaction of diverse ontogenetic and functional brain modules through development.

This view is in line with neuroconstructivist approaches to human cognition (cf. Karmiloff-Smith 2009), where language is understood primarily as a biological capacity, resulting from human-specific gene mutations affecting genes involved in brain development and wiring and language evolution, as the result of minor changes in brain wiring—although it also acknowledges a significant evolution-

ary continuity of language with the communication and cognition of other animals, in particular of non-human apes. This also results in a relative focus on the individual implementation of language, with cognitive changes resulting from neuroanatomical changes, themselves principally resulting from genetic mutations. Ultimately, supra-individual and cultural aspects of language are subordinated to individual and biological aspects, although some feedback effect of culture on cognition is expected, but not with a decisive triggering effect on language evolution. One interesting and distinctive consequence is a conception of language primarily as a tool for thinking (rather than for communicating), since our language-readiness would have initially emerged as a new, improved mechanism of conceptualising. Regarding the question of whether language is species-specific, this is mostly seen as an empirical question in these approaches, with the expectation of relatively few uniquely human components 'on top of' a majority of components shared with other species. Accordingly, this is more a gradable and gradual view of language (evolution) than a categorical one.

5. Discussion

In this paper we have argued against the intuitively appealing and occasionally expressly formulated (especially Botha 2000) dictum that language evolution as a field of research needs an overtly formulated, analytical, top-down definition of language. We have proposed that such a definition is most likely impossible, and further, even if possible, it might not be conducive to any tangible gains. As a particularly forceful illustration of this latter point, we discussed the term FLN, which was conceived in response to the perceived problem of a lack of a technical definition of (the faculty of) language, but which—as we have documented—was defined in two mutually exclusive ways, whose combining results in a tautology. In the subsequent literature the two mutually exclusive definitions widely function interchangeably, without scholars as much as noticing the conflict – a striking demonstration that the bulk of research practice in the field depends on other factors (arguably including an author's larger theoretical orientation) but not on the exact letter of the proposed top-down definitions.⁷

Notably, the claim we advance here does not imply denying the importance of clear definitions in scientific discourse generally, and in language evolution research specifically. This is no contradiction: We distinguish between definitions on two different and clearly separable levels. One is the level of more specific

⁷ Worth mentioning here is that while FLN/FLB distinction aims at a precise delineation of (the biological underpinnings of) language, other frameworks such as those that can be grouped under the umbrella of "Complex Adaptive Systems" approaches more or less explicitly acknowledge that language cannot be clearly delineated from other phenomena. In a way, these two views of language can be considered two extreme poles on the continuum from an extremely broad to a maximally narrow conceptualization. The different definitions along this continuum show that the way we think about language as an object of study partly depends on theoretical presuppositions and partly on the epistemological interest of each approach: While Chomsky, Hauser, and Fitch, for example, aim at carving out the biological, species-specific prerequisites for language and therefore narrow down the scope of language as a technical term considerably, proponents of CAS accounts take a macro-perspective on biological and cultural evolution and view language in the broader context of social-interactional phenomena.

technical terms that function as building blocks of theories and especially of hypotheses, which require unambiguous formulations so as to meet the fundamental standards of non-triviality and falsifiability. This level is thus essential for science to make progress by conclusively resolving arguments with recourse to empirical data rather than getting stuck on conceptual differences.⁸ The other level, however, is the global level of macroscopic notions, which cannot (without further specification) function as building blocks of specific theories or hypotheses but have a different role, related instead to integrative and classificatory goals.

So, for example, we largely concur with Behme that

[w]hile it may be neither feasible nor beneficial that all language evolution researchers adopt the same definition of ‘language’ it would be desirable for them to explicitly state which definition they adopt;

(2016: 8)

and with Fitch (2010: 24) that “unspecified use of [...] the word ‘language’ [...] is probably best avoided”. A small but essential caveat is that any theory-specific use of *language* will inevitably remain meronymous, in the sense of always relating only to part of the complex phenomenon. Therefore, in our view it is more productive to push definitions one level down: Leave *language* as an unanalysable prime and provide rigorous definitions of particular *components* or *aspects* of language as they function in specific theories under consideration. One very important advantage is that such a strategy prevents attempts to monopolise the word *language* by a particular theory that would claim unique privileged access to a ‘correct’ understanding of *language*, something that Chomskyan approaches have been criticised for.

Exactly such was the nature of the FLN/FLB distinction (again, two different distinctions, as we show in Section 3), categorical about the nature of language and expressly formulated to guide language evolution research as an understanding of the language faculty privileged over other theories. Interestingly, however, much fuzzier notions of language seem to have better served the actual language evolution research. Due to the breadth of research interests in language evolution in the last decade, and the intense interdisciplinarity that cuts across many disparate areas of investigation—from computational modelling, to primate communication, to sign linguistics—language evolution thrives on fuzzy definitions of language and finds categorical, top-down approaches too constraining. This idea is reflected in Section 4, which surveys influential lines of research in language evolution, mostly focusing on the most recent trends. In recent years, the bulk of research in this field revolves around the problems of multimodality, the dynamics of cultural transmission, language as a form of social interaction or biological language-readiness. What emerges from this survey is indeed a breadth of the range of these perspectives that precludes their fitting together under any single definition of language.

⁸ We are grateful to two anonymous reviewers for comments that led us to stress this important point.

This is far from surprising if we look to analogies in other notions in science that, like language, are macroscopic, unobservable and unoperationalisable. One example comes from biologists, who tend to avoid top-down approaches to life (and aprioristic definitions of *life*) and focus instead on the study of the building blocks of living beings, regardless of whether they can be found in other domains (e.g., water) or not (the DNA). Only a successful characterization of these building blocks can lead to achieving a comprehensive view of the nature of life (as in systems biology) and its evolution (as in, e.g., evolutionary developmental biology). Similarly, analyses of the use of terms such as *heat* in physics (Lewis & Linn 1996) or *gesture* in primatology (Bourjade et al. 2020) provide arguments for a beneficial and productive role of conceptual diversity, at least when certain conditions are met such as consistent use of a term within a particular approach.

One particularly interesting motivation for why a lack of a single top-down definition of language in language evolution is not as consequential now as it was 20 years ago may be a methodological change in the profile of this field, from theoretical to empirical research (see especially Dediu & de Boer 2016, Fitch 2017, Żywicznyński 2018, Nölle et al. 2020). In 2017, Waciewicz & Żywicznyński wrote:

Language evolution researchers no longer stop at being consumers of empirical data, but rather aim at being providers as well, acquiring data by experimentation, observation, or simulation (and a steadily increasing proportion of these results then feed back into more general discussions on the nature of language [...]). The maturation of language evolution research has been marked by a steady growth in the proportion of empirical (“new data”) research relative to theoretical (synthetic) argumentation [...]. In the volume that grew out of the first EVOLANG conference in 1996 (Hurford et al. 1998), all 24 contributions have a decidedly theoretical (synthesising) character, whereas the proceedings of the most recent conference (Roberts et al. 2016) are dominated by empirical research: 123 contributions, as opposed to 25 theoretical.

Waciewicz & Żywicznyński (2017: 3)

What follows is that recently, very few publications present comprehensive scenarios of language evolution, and conversely, a vast proportion of studies are more fine-grained, addressing much more specific and narrower Kuhnian “puzzles” such as the efficiency of gestural vs multimodal signals in conveying emotional meanings (Zlatev et al. 2017) or the effect of processed food on the dental configuration and in turn on the production of fricatives (Blasi et al. 2019). A natural consequence is that such specific and bottom-up studies do not directly aim at explaining language evolution *sensu largo*, and so do not need to work with a definition of language *sensu largo*.

An epitome of both the empirical and bottom-up approach is the Causal Hypotheses In Evolutionary Linguistics Database (CHIELD, pronounced ‘shield’; Roberts et al. 2020). CHIELD contains crowd-sourced entries for over 400 publications, with over 3,400 causal links between more than 1,700 variables and aims not only at cataloguing hypotheses about language evolution but also making data on them interoperable. As it is unlikely that all 32 authors (much less all 41

contributors to the database) would converge on a single explicit definition of language, CHIELD is possible not *despite* but *because* it deliberately avoids defining language (cf. “A classic example of this is the word ‘language’ itself, which can be interpreted as anything relating to human communication or only a specific syntactic ability”; Roberts et al. 2020: 3).

Consequently, one way of describing language evolution could be to ‘bracket’ the notion of *language* and rely solely on content-independent, institutional and scientometric criteria such as conferences, journals, laboratories and citation patterns (cf. Bergmann & Dale 2016). This would delineate a collection of bottom-up approaches and researchers that jointly form a ‘community of practice’ or a *denkkolektivo* (Fleck 1979). This is an interesting approach with some genuine explanatory power; for example, this strategy would address Haspelmath’s (2016) question of why *Journal of Language Evolution* publishes research on language change of apparently non-evolutionary character.

Nevertheless, such an approach would seem deeply unsatisfying to the researchers in the field of language evolution, who have a strong sense of unifying research substance, and in particular the unifying aim of explaining the origin of language. This substance is primary to the content-independent factors, in that it provides identity to the field and gives rise to—as opposed to being secondary and merely resultant from—the patterns and networks of personal and institutional connections. It is the basic, common, intuitive understanding of *language*, and basic human curiosity about how it began, that sets the explanatory goal for the field of language evolution as a whole, and thus shapes its research practices and the resulting *denkkollektivo*—rather than vice versa.

This is the other point that follows from our review: although the recently most influential approaches to language evolution are indeed too diverse to be brought together under a common definition of language, they also do overlap to a large extent in terms of key definitional dimensions. As discussed in section 4, these dimensions are in particular the criterial components of language, its modalities, domain-specificity, biological versus cultural profile, (supra)individual character, gradability, species-specificity and primary function. This complex pattern of numerous similarities and sporadic but significant differences is characteristic of a family-resemblance category (Wittgenstein 1953). Most importantly from the point of view of research practice, this family resemblance pattern underwrites fruitful communication between these approaches, leading to cross-fertilisation and opening new research vistas: for example, there is a growing number of studies in the Iterated Learning paradigm that look at different communicative modalities (e.g., Motamedi et al. 2019); as another example, the self-domestication theory, proposed within the Complex Adaptive Systems camp, is now most actively developed by the proponents of the language ready brain (see especially Benítez-Burraco & Progovac 2020). Thus, based on a tacit and fundamentally ineffable notion of language, all these approaches jointly contribute to the development of the field of language evolution.

Acknowledgements

This research was supported by the Polish National Science Centre under grant agreement UMO-2019/34/E/HS2/00248. This work was also supported from the Excellence Initiative—Research University at Nicolaus Copernicus University in Toruń. We are grateful to Marta Sibierska for her valuable comments and to Aleksandra Poniewierska and Aleksandra Szczepańska for their assistance in editing the manuscript.

References

- Aboitiz, Francisco. 2012. Gestures, vocalizations, and memory in language origins. *Frontiers in Evolutionary Neuroscience* 4(2).
- Amici, Federica, Alex Sánchez-Amaro, Carla Sebastián-Enesco, Trix Cacchione, Matthias Allritz, Juan Salazar-Bonet, & Federico Rossano. 2019. The word order of languages predicts native speakers' working memory. *Scientific Reports* 9(1), 1124.
- Arbib, Michael A. 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28(2), 105–167.
- Arbib, Michael A. 2012. *How the Brain Got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press.
- Arbib, Michael A. 2016. Towards a computational comparative neuroprimatology: framing the language-ready brain. *Physics of Life Reviews* 16, 1–54.
- Arbib, Michael A. 2018. Computational challenges of evolving the language-ready brain: 2. Building towards neurolinguistics. *Interaction Studies* 19(1–2), 22–37.
- Arbib, Michael A, Francisco Aboitiz, Judith M. Burkart, Michale C. Corballis, Gino Coudé, Erin Hecht, Katja Liebal, Masako Myowa-Yamakoshi, James Pustejovsky, Shelby S. Putt, Federico Rossano, Anne E. Russon, P. Thomas Schoenemann, Uwe Seifert, Katerina Semendeferi, Chris Sinha, Dietrich Stout, Virginia Volterra, Sławomir Waciewicz, & Benjamin Wilson. 2018. The comparative neuroprimatology 2018 (CNP-2018) road map for research on how the brain got language. *Interaction Studies* 19, 371–388.
- Armstrong, David F. & Sherman E. Wilcox. 2007. *The Gestural Origin of Language*. Oxford: Oxford University Press.
- Balari, Sergio & Guillermo Lorenzo. 2012. *Computational Phenotypes: Towards an Evolutionary Developmental Biolinguistics*. Oxford: Oxford University Press.
- Balari, Sergio & Guillermo Lorenzo. 2016. Evo-devo of language and cognition. In Laura Nuno de la Rosa & Gerd Müller (eds.), *Evolutionary Developmental Biology*. Cham: Springer.
- Barceló-Coblijn, Lluís. 2012. Evolutionary scenarios for the emergence of recursion. *Theoria et Historia Scientiarum* 9, 171–199.
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59, 1–26.

- Behme, Christina. 2016. Evolution of what? In Sean G. Roberts, Christine Cuskley, Luke McCrohon, Lluís Barceló-Coblijn, Olga Fehér, & Tessa Verhoef (eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVLANG11)*. <http://evolang.org/neworleans/papers/140.html>
- Benítez-Burraco, Antonio. 2020. Prehistoric languages and human self-domestication. *Language Dynamics and Change* 10(1), 27–58.
- Benítez-Burraco, Antonio & Cedric Boeckx. 2014. Universal Grammar and biological variation: An evodevo agenda for comparative biolinguistics. *Biological Theory* 9(2), 122–134.
- Benítez-Burraco, Antonio & Cedric Boeckx. 2015. Possible functional links among brain- and skull-related genes selected in modern humans. *Frontiers in Psychology* 6: 794.
- Benítez-Burraco, Antonio & Ljiljana Progovac. 2020. A four-stage model for language evolution under the effects of human self-domestication. *Language & Communication* 73, 1–17.
- Bergen, Benjamin K. 2012. *Louder than Words: The New Science of How the Mind Makes Meaning*. New York: Basic Books.
- Bergmann, Till & Rick Dale. 2016. A scientometric analysis of evolang: Intersections and authorships. In Sean G. Roberts, Christine Cuskley, Luke McCrohon, Lluís Barceló-Coblijn, Olga Fehér, & Tessa Verhoef (eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVLANG11)*. <http://evolang.org/neworleans/papers/182.html>
- Berwick, Robert C. & Noam Chomsky. 2016. *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press.
- Bickerton, Derek. 1990. *Language and Species*. Chicago, IL: University of Chicago Press.
- Blasi, Damián E., Steven Moran, Scott R. Moisiuk, Paul Widmer, Dan Dediu, & Balthasar Bickel. 2019. Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science* 363(6432).
- Boeckx, Cedric. 2012. The I-language mosaic. In Cedric Boeckx, María del Carmen Horno Chéliz, & José Luis Mendívil Giró (eds.), *Language, from a Biological Point of View: Current Issues in Biolinguistics*, 23–51. Newcastle upon Tyne: Cambridge Scholars.
- Boeckx, Cedric & Antonio Benítez-Burraco. 2014a. Globularity and language-readiness: Generating new predictions by expanding the set of genes of interest. *Frontiers in Psychology* 5(1324), doi: 10.3389/fpsyg.2014.01324.
- Boeckx, Cedric & Antonio Benítez-Burraco. 2014b. The shape of the human language-ready brain. *Frontiers in Psychology* 5(282), doi: 10.3389/fpsyg.2014.00282.
- Bolhuis, Johan J., Ian Tattersall, Noam Chomsky & Robert C. Berwick. 2014. How could language have evolved? *PLoS Biology* 12(8), doi: 10.1371/journal.pbio.1001934.
- Botha, Rudolf P. 2000. Discussing the evolution of the assorted beasts called language. *Language & Communication* 20(2), 149–160.
- Bouchard, Denis. 2013. *The nature and origin of language*. Oxford: Oxford University Press.

- Bouchard, Denis. 2015. Brain readiness and the nature of language. *Frontiers in Psychology* 6. doi:10.3389/fpsyg.2015.01376
- Bourjade, Marie, Hélène Cochet, Sandra Molesti & Michèle Guidetti. 2020. Is conceptual diversity an advantage for scientific inquiry? A case study on the concept of 'gesture' in comparative psychology. *Integrative Psychological and Behavioral Science*, doi: 10.1007/s12124-020-09516-5.
- Bowling, Daniel L. 2017. The continuing legacy of nature versus nurture in biolinguistics. *Psychonomic Bulletin & Review* 24(1). 140–141.
- Bruner, Emiliano, Giorgio Manzi & Juan L. Arsuaga. 2003. Encephalization and allometric trajectories in the genus *Homo*: evidence from the Neandertal and modern lineages. *Proceedings of the National Academy of Sciences* 100. 15335–15340.
- Byrne, Richard & Andrew Whiten (eds.). 1988. *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford: Oxford University Press.
- Cheney, Dorothy L. & Robert M. Seyfarth. 2007. *Baboon metaphysics: The evolution of a social mind*. Chicago: Chicago University Press.
- Christiansen, Morten H. & Simon Kirby. 2003. Language evolution: Consensus and controversies. *Trends in Cognitive Sciences* 7(7). 300–307.
- Christiansen, Morten H. & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31(5). 489–509.
- Corballis, Michael C. 2013. Gestural theory of the origins of language. In Claire Lefebvre, Bernard Comrie & Henri Cohen (eds.), *New perspectives on the origins of language*, 171–184. Oxford: Oxford University Press.
- Cornish, Hannah. 2010. Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems. *Interaction Studies* 11(1). 112–137.
- Cowie, Fiona. 2008. Innateness and language. *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2017/entries/innateness-language/>. (21 July, 2020.)
- Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Pearson Longman.
- Croft, William. 2011. Language structure in its human context: New directions for the language sciences in the twenty-first century. In Patrick Colm Hogan (ed.), *The Cambridge encyclopedia of the language sciences*, 1–11. Cambridge: Cambridge University Press.
- Darwin, Charles. 1871. *The descent of man: and selection in relation to sex*. London: J. Murray.
- Deacon, Terrence W. 1997. *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York & London: Norton.
- de Boer, Bart. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28. 441–465.
- Dediu, Dan & Bart de Boer. 2016. Language evolution needs its own journal. *Journal of Language Evolution* 1. 1–6.
- Divjak, Dagmar. 2019. *Frequency in language: memory, attention and learning*. Cambridge: Cambridge University Press.

- Donald, Merlin. 1991. *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.
- Donald, Merlin. 2001. *A mind so rare: The evolution of human consciousness*. New York: Norton.
- Dor, Daniel. 2015. *The Instruction of Imagination: language as a social communication technology*. Oxford: Oxford University Press.
- Dor, Daniel, Chris Knight, Jerome Lewis (eds.). 2014. *The social origins of language*. Oxford: Oxford University Press.
- Dunbar, Robin I. M. 1993. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*. 16(04). 681–694.
- Dunbar, Robin I. M. 1996. *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.
- Ejiri, Keiko & Nobuo Masataka. 2001. Co-occurrences of preverbal vocal behavior and motor action in early infancy. *Developmental Science* 4(1). 40–48.
- Ekman, Paul & Wallace V. Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1(1). 49–98.
- Elman, Jeffrey L. 1999. The emergence of language: A conspiracy theory. In Brian MacWhinney (ed.), *The Emergence of Language*. 1–28. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5). 429–448.
- Fitch, Tecumseh W. 2010. *The evolution of language*. Cambridge: Cambridge University Press.
- Fitch, Tecumseh W. 2017. Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review* 24(1). 3–33.
- Fitch, Tecumseh W., Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: clarifications and implications. *Cognition* 97(2). 179–210.
- Fleck, Ludwik. 1979. *Genesis and Development of a Scientific Fact*. Chicago: The University of Chicago Press.
- Garrod, Simon, Nicolas Fay, Shane Rogers, Bradley Walker & Nik Swoboda. 2010. Can iterated learning explain the emergence of graphical symbols? *Interaction Studies* 11(1). 33–50.
- Green, Jennifer. 2014. *Drawn from the ground: Sound, sign and inscription in Central Australian sand stories*. Cambridge: Cambridge University Press.
- Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*, 41–58. New York: Academic Press.
- Gruberger, Michal, Eti Ben-Simon, Yechiel Levkovitz, Abraham Zangen & Talma Hendler. 2011. Towards a neuroscience of mind-wandering. *Frontiers in Human Neuroscience* 5(56), doi: 10.3389/fnhum.2011.00056.
- Gunz, Philipp, Simon Neubauer, Bruno Maureille & Jean-Jacques Hublin. 2010. Brain development after birth differs between neanderthals and modern humans. *Current Biology* 20. R921–R922.
- Gunz, Philipp, Simon Neubauer, Liubov V. Golovanova, Vladimir Doronichev, Bruno Maureille & Jean-Jacques Hublin. 2012. A uniquely modern human pattern of endocranial development. insights from a new cranial

- reconstruction of the Neandertal newborn from Mezmaiskaya. *Journal of Human Evolution* 62. 300–313.
- Haspelmath, Martin. 2016. The evolution (or diachrony) of “language evolution”. <http://dlc.hypotheses.org/894>. (21 July, 2020.)
- Haspelmath, Martin. 2020. Human linguisticity and the building blocks of languages. *Frontiers in Psychology* 10. 3056. doi: 10.3389/fpsyg.2019.03056.
- Hauser, Marc, Noam Chomsky & Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298. 1569–1579.
- Hauser, Marc, David Barner & Tim O’Donnell. 2007. Evolutionary linguistics: A new look at an old landscape. *Language Learning and Development* 3(2). 101–132.
- Hecht, Erin E., David A. Gutman, Todd M. Preuss, Mar M. Sánchez, Lisa A. Parr & James K. Rilling. 2013. Process versus product in social learning: comparative diffusion tensor imaging of neural systems for action execution-observation matching in macaques, chimpanzees, and humans. *Cortex* 23. 1014–1024.
- Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press.
- Hewes, Gordon Winant. 1977. Language origin theories. In Duane M. Rumbaugh (ed.), *Language learning by a chimpanzee: The Lana project*, 3–53. New York: Academic Press.
- Hrdy, Sarah Blaffer. 2009. *The woman that never evolved*. Cambridge, MA: Harvard University Press.
- Hurford, James R. 1989. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* 77(2). 187–222.
- Hurford, James R. 2003. The language mosaic and its evolution. In Morten H. Christiansen & Simon Kirby (eds.), *Language Evolution* 3, 38–57. Oxford: Oxford University Press.
- Hurford, James R. 2007. *The Origins of Meaning: Language in the Light of Evolution, Vol. 1*. Oxford: Oxford University Press.
- Hurford, James R. 2012. *The Origins of Grammar: Language in the Light of Evolution, Vol. 2*. Oxford: Oxford University Press.
- Hurford, James R., Michael Studdert-Kennedy & Chris Knight (eds.). 1998. *Approaches to the evolution of language: social and cognitive bases*. Cambridge: Cambridge University Press.
- Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Jackendoff, Ray. 2010. Your theory of language evolution depends on your theory of language. in Richard K. Larson, Viviane Déprez & Hiroko Yamakido (eds), *The evolution of human language: Bilingual perspectives*, 63–72. Cambridge: Cambridge University Press.
- Jackendoff, Ray & Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97(2). 211–225.
- Johansson, Sverker. 2005. *Origins of language: Constraints on hypotheses*. John Benjamins Publishing.

- Karmiloff-Smith, Annette. 2009. Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Neurocognition* 45. 56–63.
- Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key (ed.), *The Relationship of verbal and nonverbal communication*, 207–227. The Hague & New York: Mouton.
- Kendon, Adam. 1990. *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.
- Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kendon, Adam. 2008. Signs for language origins? *The Public Journal of Semiotics* 2(2). 2–29.
- Kendon, Adam. 2011. Some modern considerations for thinking about language evolution: A discussion of the evolution of language by Tecumseh Fitch. *The Public Journal of Semiotics* 3(1). 79–108.
- Kendon, Adam. 2014a. Semiotic diversity in utterance production and the concept of ‘language.’ *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). doi:10.1098/rstb.2013.0293
- Kendon, Adam. 2014b. The “poly-modalic” nature of utterances and its relevance. In Daniel Dor, Chris Knight & Jerome Lewis (eds.), *The social origins of language*, 67–76. Oxford: Oxford University Press.
- Kendon, Adam. 2017. Reflections on the “gesture-first” hypothesis of language origins. *Psychonomic Bulletin & Review* 24(1). 163–170.
- Kinsella, Anna R. 2009. *Language evolution and syntactic theory*. Cambridge: Cambridge University Press.
- Kirby, Simon. 2012. Language is an adaptive system: The role of cultural evolution in the origins of structure. In Maggie Tallerman & Kathleen R. Gibson (eds.), *The Oxford Handbook of Language Evolution*, 589–604. Oxford: Oxford University Press.
- Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31). 10681–10686.
- Kirby, Simon, Mike Dowman & Thomas L. Griffiths. 2007. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences* 104(12), 5241–5245.
- Kirby, Simon, Tom Griffiths & Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* 28. 108–114.
- Kurcz, Ida. 2004. Communicative competence and theory of mind. *Psychology of Language and Communication* 8(2). 5–18.
- Larsen-Freeman, Diane. 2017. Complexity theory: The lessons continue. In Lourdes Ortega & ZhaoHong Han (eds.), *Complexity Theory and Language Development: In celebration of Diane Larsen-Freeman*, 11–50. Amsterdam, Philadelphia: John Benjamins.
- Levinson, Stephen C. 2006. On the human “interaction engine”. In Nick J. Enfield & Stephen C. Levinson (eds.), *Roots of Human Sociality*. 39–69. Oxford: Berg.

- Levinson, Stephen C., & Judith Holler. 2014. The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). 20130302.
- Lewandowska-Tomaszczyk, Barbara. 2008. Czym jest język? Dzisiejsze kontrowersje w paradygmatach generatywnych i kognitywnych. In Piotr Stalmaszczyk (ed.), *Metodologie językoznawstwa: współczesne tendencje i kontrowersje*, 9–26. Kraków: Lexis.
- Lewis, Eileen L. & Marcia C. Linn. 1996. Where is the heat? A response to David Pushkin. *Journal of Research in Science Teaching* 33. 335–337.
- Lohmann, Heidemarie & Michael Tomasello. 2003. The role of language in the development of false belief understanding: A training study. *Child Development* 74(4). 1130–1144.
- Little, Hannah, Kerem Eryilmaz & Bart de Boer 2017. Signal dimensionality and the emergence of combinatorial structure. *Cognition* 168. 1–15.
- Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS One* 5. e8559.
- MacNeilage, Peter F. 2008. *The origin of speech*. Oxford & New York: Oxford University Press.
- Martins, Pedro Tiago, Evelina Leivada, Antonio Benítez-Burraco & Cedric Boeckx. 2016. Biological pluralism in service of biolinguistics. In Koji Fujita & Cedric Boeckx (eds.), *Advances in biolinguistics: The human language faculty and its biological basis*, 153–169. New York: Routledge.
- McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, David. 2012. *How language began: Gesture and speech in human evolution*. Cambridge: Cambridge University Press.
- Mead, George H. 1974. *Mind, Self and Society*, edn. with introduction by Charles W. Morris. Chicago: University of Chicago Press.
- Mendívil-Giró, José-Luis. 2019. Did language evolve through language change? On language change, language evolution and grammaticalization theory. *Glossa* 4(1), <https://www.glossa-journal.org/articles/10.5334/gjgl.895/>.
- Mesoudi, Alex. 2011. *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences*. University of Chicago Press.
- Miller, Earl K. & Timothy J. Buschman. 2013. Cortical circuits for the control of attention. *Current Opinion in Neurobiology* 23. 216–222.
- Motamedi, Yasamin, Marieke Schouwstra, Kenny Smith, Jennifer Culbertson & Simon Kirby. 2019. Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition* 192. 103964.
- Neubauer, Simon, Philipp Gunz, Jean-Jacques Hublin. 2010. Endocranial shape changes during growth in chimpanzees and humans: a morphometric analysis of unique and shared aspects. *Journal of Human Evolution* 59. 555–566.
- Nölle, Jonas, Stefan Hartmann & Peter Tinitis. 2020. Language evolution research in the year 2020. *Language Dynamics and Change* 10(1). 3–26.
- Okanoya, Kazuo. 2007. Language evolution and an emergent property. *Current Opinion in Neurobiology* 17(2). 271–276.
- Parker, Anna R. 2006. Evolving the narrow language faculty: was recursion the pivotal step? *The Evolution of Language*. 239–246.

- Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: what's special about it? *Cognition* 95(2). 201–236.
- Pleyer, Michael & James Winters. 2014. Integrating cognitive linguistics and language evolution research. *Theoria et Historia Scientiarum* 11. 19–43.
- Pleyer, Michael & Stefan Hartmann. 2019. Constructing a consensus on language evolution? Convergences and differences between biolinguistic and usage-based approaches. *Frontiers in Psychology: Language Sciences*. doi:10.3389/fpsyg.2019.02537
- Poeppl, David. 2012. The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology* 29. 34–55.
- Poeppl, David & David Embick. 2005. Defining the relation between linguistics and neuroscience. In Anne Cutler (ed.), *Twenty-first Century Psycho-linguistics: Four Cornerstones*, 103–120. Hillsdale: Lawrence Erlbaum.
- Pomerantz, Anita. 1984. Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In Maxwell Atkinson & John Heritage (eds.), *Structures of Social Action: Studies in Conversation Analysis*, 57–101. Cambridge: Cambridge University Press.
- Richerson, Peter J. & Robert Boyd. 2005. *Not by genes alone: how culture transformed human evolution*. Chicago: University of Chicago Press.
- Ritt, Nikolaus. 2004. *Selfish Sounds and Linguistic Evolution: A Darwinian Approach to Language Change*. Cambridge: Cambridge University Press.
- Roberts, Seán G., Christine Cuskley, Luke McCrohon, Lluís Barceló-Coblijn, Olga Fehér & Tessa Verhoef (eds.). 2016. *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*. doi:10.17617/2.2248195
- Roberts, Seán G., Anton Killin, Angarika Deb, Catherine Sheard, Simon J. Greenhill, Kaius Sinnemäki, José Segovia-Martín, Jonas Nölle, Aleksandrs Berdicevskis, Archie Humphreys-Balkwill, Hannah Little, Christopher Opie, Guillaume Jacques, Lindell Bromham, Peeter Tinitis, Robert M. Ross, Sean Lee, Emily Gasser, Jasmine Calladine, Matthew Spike, Stephen Francis Mann, Olena Shcherbakova, Ruth Singer, Shuya Zhang, Antonio Benítez-Burraco, Christian Kliesch, Ewan Thomas-Colquhoun, Hedvig Skirgård, Monica Tamariz, Sam Passmore, Thomas Pellard & Fiona Jordan. 2020. CHIELD: the causal hypotheses in evolutionary linguistics database. *Journal of Language Evolution*. doi:10.1093/jole/lzaa001
- Samuels, Bridget. 2009. The third factor in phonology. *Biolinguistics* 3(2–3). 355–382.
- Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris: Payot.
- Scott-Phillips, Thomas C. 2015. *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Houndmills: Palgrave Macmillan.
- Scott-Phillips, Thomas C. & Simon Kirby. 2010. Language evolution in the laboratory. *Trends in cognitive sciences* 14(9). 411–417.
- Smith, Kenny, Simon Kirby & Henry Brighton. 2003. Iterated learning: A framework for the emergence of language. *Artificial Life* 9. 371–386.
- Smith, Kenny. 2018. The cognitive prerequisites for language: insights from iterated learning. *Current Opinion in Behavioral Sciences* 21. 154–160.

- Smith, Maynard John & Eörs Szathmáry. 1995. *The major transitions in evolution*. Oxford & New York: W.H. Freeman Spektrum.
- Smith, Kenny & Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116(3). 444–449.
- Sonesson, Göran. 1997. The ecological foundations of iconicity. In Irmengard Rauch & Gerald Carr (eds.), *Semiotics Around the World: Synthesis in Diversity*, 739–742. Berlin & New York: Mouton de Gruyter.
- Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
- Steels, Luc. 2000. Language as a complex adaptive system. In Marc Schoenauer, Kalyanmoy Deb, Günter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo & Hans-Paul Schwefel (eds.), *Parallel Problem Solving from Nature*, 17–26. Berlin, Heidelberg: Springer.
- Számadó, Szabolcs & Eörs Szathmáry. 2006. Selective scenarios for the emergence of natural language. *Trends in Ecology & Evolution* 21(10). 555–561.
- Tallerman, Maggie & Kathleen R. Gibson (eds.). 2012. *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press.
- Thomas, James & Simon Kirby. 2018. Self domestication and the evolution of language. *Biology & Philosophy* 33(1). 9.
- Tincoff, Ruth & Marc D. Hauser. 2006. Cognitive basis for language evolution in nonhuman primates. In Keith Brown (ed.), *The encyclopedia of language & linguistics*, 533–538. 2nd edn. Amsterdam: Elsevier.
- Tinits, Peter, Jonas Nölle & Stefan Hartmann. 2017. Usage context influences the evolution of overspecification in iterated learning. *Journal of Language Evolution* 2(2). 148–159.
- Tomasello, Michael. 1999. *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, Michael. 2003. On the different origins of symbols and grammar. In Morten H. Christiansen & Simon Kirby (eds.), *Language Evolution*, 94–110. Oxford University Press.
- Tomasello, Michael. 2006. Why don't apes point? In Nick J. Enfield & Stephen C. Levinson (eds.), *Roots of human sociality: Culture, cognition and interaction*, 506–524. Oxford: Berg.
- Tomasello, Michael. 2008. *Origins of human communication*. Cambridge, MA: MIT Press.
- Tomasello, Michael. 2009. The usage-based theory of language acquisition. In Edith Laura Bavin (ed.), *The Cambridge Handbook of Child Language*, 69–87. Cambridge: Cambridge University Press.
- Tomasello, Michael. 2019. *Becoming human: A theory of ontogeny*. Cambridge, MA: Belknap Press.
- Tomasello, Michael & Malinda Carpenter. 2007. Shared intentionality. *Developmental Science* 10(1). 121–125.
- Tomasello, Michael, Ann Cale Kruger & Hilary Horn Ratner. 1993. Cultural learning. *Behavioral and Brain Sciences* 16(3). 495–511.
- Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne & Henrike Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28(5). 675–691.

- Trijp, Remi van. 2011. Can iterated learning explain the emergence of case marking in language? *Proceedings of the 23rd Benelux Conference on Artificial Intelligence (BNAIC 2011), Volume 1*. 288–295.
- Tuomela, Raimo. 2007. *The philosophy of sociality: The shared point of view*. Oxford: Oxford University Press.
- Verhoef, Tessa, Simon Kirby & Bart de Boer. 2014. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics* 43. 57–68.
- Vigliocco, Gabriella, Pamela Perniss & David Vinson. 2014. Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). 1–7. <https://royalsocietypublishing.org/doi/10.1098/rstb.2013.0292>. (24 July, 2020.)
- Wacewicz, Sławomir. 2007. Debata Hauser, Chomsky, Fitch – Pinker i Jackendoff. Nowoczesny spór o pochodzenie ludzkiej zdolności językowej. In Szymon Wróbel (ed.), *Modularność umysłu*, 79–94. Kalisz: WPA UAM.
- Wacewicz, Sławomir. 2012. The narrow faculty of language: What is it, who has it, and how is it defined? *Theoria et Historia Scientiarum* 9. 217–229.
- Wacewicz, Sławomir & Przemysław Żywiczyński. 2017. The multimodal origins of linguistic communication. *Language & Communication* 54. 1–8.
- Wescott, Roger W. 1991. Defining Language. In Walburga von Raffler-Engel, Jan Wind & Abraham Jonker (eds.), *Studies in Language Origins: Volume 2*. John Benjamins Publishing. 77–84.
- Willems, Roel M., Asli Özyürek & Peter Hagoort. 2007. When language meets action: The neural integration of gesture and speech. *Cerebral Cortex* 17(10). 2322–2333.
- Wittgenstein, Ludwig 1953. *Philosophical Investigations*. Translated by Elisabeth Anscombe. Oxford: Basil Blackwell.
- Wróbel, Szymon. 2012. Rethinking language faculty: Has language evolved for other than language related reasons? *Theoria et Historia Scientiarum* 9, 201–216.
- Zlatev, Jordan. 2008. From proto-mimesis to language: Evidence from primatology and social neuroscience. *Journal of Physiology-Paris* 102(1–3), 137–151.
- Zlatev, Jordan. 2014. Human uniqueness, bodily mimesis and the evolution of language. *Humana. Mente Journal of Philosophical Studies* 7(27), 197–219.
- Zlatev, Jordan. 2016. Preconditions in human embodiment for the evolution of symbolic communication In Gregor Etzelmüller & Christian Tewes (eds.), *Embodiment in Evolution and Culture*, 151–174. Tübingen: Mohr Siebeck.
- Zlatev, Jordan. 2019. Mimesis theory, learning and polysemiotic communication. In Michael Peters (ed.), *Encyclopedia of Educational Philosophy and Theory*. Dordrecht: Springer.
- Zlatev, Jordan, Sławomir Wacewicz, Przemysław Żywiczyński, & Joost van de Weijer. 2017. Multimodal-first or pantomime-first? Communicating events through pantomime with and without vocalization. *Interaction Studies* 18(3), 465–488.

- Zlatev, Jordan, Przemysław Żywiczyński & Sławomir Wacewicz. 2020. Pantomime as the original human-specific communication system. *Journal of Language Evolution* 5(2), 156–174. doi:10.1093/jole/lzaa006
- Żywiczyński, Przemysław. 2018. *Language Origins: From Mythology to Science*. Berlin: Peter Lang.
- Żywiczyński, Przemysław & Jordan Zlatev. In press. The role of gesture in debates on the origins of language. In Alan Cienki (ed.), *Handbook of Gesture Studies*. Cambridge: Cambridge University Press.

The Phonological Latching Network

Joe Stephen Bratsvedal Collins^{1,2,3,*}

¹ The Center for Advanced Study in Theoretical Linguistics (CASTL), UiT – The Arctic University of Norway, Tromsø, Norway

² Cognitive Neuroscience Group, International School for Advanced Studies (SISSA), Trieste, Italy

³ Department of Language and Literature, Faculty of Humanities, Norwegian University of Science and Technology, Trondheim, Norway

* Correspondence: joe.collins@ntnu.no

 JSBC: <https://orcid.org/0000-0001-7335-5134>

This paper gives an analysis of an attractor neural network model dubbed the Phonological Latching Network. The model appears to reproduce certain quintessentially phonological phenomena, despite not having any of these phonological behaviours programmed or taught to the model. Rather, assimilation, segmental-OCP, and sonority sequencing appear to emerge spontaneously from the combination of a few basic brain-like ingredients with a phonology-like feature system. The significance of this can be interpreted from two angles: firstly, the fact that the model spontaneously produces attested natural language patterns can be taken as evidence of the model's neural and psychological plausibility; and secondly, it provides a potential explanation for why these patterns appear to frequently in natural language grammars. Namely, they are a consequence of latching dynamics in the brain.

Keywords: phonology; neuroscience; neural networks; attractors; Potts model


1. Introduction

In 1887, Albert Fournie claimed that “[s]peech is the only window through which the physiologist can view the cerebral life” (translation from Lashley 1951). There is nothing novel then, in the claim that the study of language should provide some insight into the workings of the human mind/brain. Indeed, even today, this is one of few mantras shared by linguists of the seemingly irreconcilable “Gener-

Editors: Lluís Barceló-Coblijn, Universitat de les Illes Balears, Spain
Evelina Leivada, Universitat Rovira i Virgili, Spain

Received: 27 December 2019
Accepted: 19 November 2020
Published: 25 March 2021

ISSN 1450–3417

 CC BY 4.0 License
© 2020 The authors

ative” and “Cognitive” schools (e.g., Chomsky 2002; Lakoff 1988). Given this apparent consensus then, it is perhaps surprising that no breakthrough in our understanding of the brain can yet be attributed to some insight from the study of language.

An analysis and critique of this state of affairs is given by Poeppel & Embick (2005), who identify (amongst other things) that we currently have no way of relating the ontologies of linguistics and neuroscience. This *Ontological Incommensurability Problem* (OIP) can be resolved, they argue, by the use of a *Linking Hypothesis*, which spells out linguistic computations at the relevant level of algorithmic abstraction, such that the neuroscientist need only find the exact implementations of those algorithms in the brain. If such a hypothesis were sufficiently complete then it could, in principle, predict the kinds of neural configurations required for natural language processing, using linguistic theories as their starting point. In this way, we could finally realize the long sought-after goal of cashing in theories of language for understanding of the human brain. Simultaneously, a *Linking Hypothesis* also has the potential to unearth lower-level explanations for linguistic phenomena, for example where those explanations might depend on purely neurobiological notions (e.g., neuronal morphology, synaptic density, metabolic efficiency, etc.).

1.1. *Emergence as a Linking Hypothesis*

The specific approach to the OIP advocated by Poeppel & Embick treats the neurobiological level of analysis as something akin to a decomposition of a linguistic theory. That is, a linguistic theory can be reduced to individual processes (e.g., concatenation, linearization, etc.), and the problem of how to realise each process can be attacked individually. And, while this approach is certainly a logical possibility for resolving the OIP, it rests on assumptions which treat the brain as being fundamentally like a digital, programmable computer. Implicitly, it has borrowed from computer science the idea that the different levels of abstraction for which we might describe a cognitive function, are related to one another through a strict compositional semantics. That is, any property at one level of abstraction can be neatly decomposed to some combination of properties at a lower level of abstraction (e.g., Block 1995).

A full rebuttal of these assumptions is well beyond the scope of this article. It is sufficient to note that this view is by no means the only starting point for constructing a *Linking Hypothesis*. The alternate approach offered here draws inspiration from the natural sciences, where the apparent incommensurability between different levels of abstraction is frequently resolved by treating the higher levels as *epistemologically emergent*¹ from lower ones (e.g., Anderson 1972; Luisi 2002). According to this approach, the goal is not to decompose a macro-level ontology to see how each component is “implemented” at the micro-level.

¹ Alternatively: *weakly emergent* (Bedau 1997). Also note that this notion of *emergence* is strictly orthogonal to the notion of *ontogenetic emergence* employed in the study of language acquisition. Whether linguistic ontology is *epistemologically/weakly emergent* does not predict whether it is learned/innate/none of the above.

Rather, the goal is to see what kinds of configurations at the micro-level give rise to a complex system whose behaviour is captured by the macro-level theory.

Therefore, to claim that linguistics is *emergent* from neuroscience entails that linguistic properties do not separately decompose to neuroscientific properties, contra the way that the functions of a high-level computer language reduce to combinations of primitive operations. Instead, the relationship between linguistics and neuroscience would be analogous to, for example, the molecular theory of gasses.² Under this view, linguistic properties would be analogous to macro-level concepts like *temperature* or *pressure*, while neuroscientific properties are analogous to molecular explanations of these phenomena. The most relevant aspect of this analogy is that the properties present at each level of abstraction are quite different. So different, in fact, that the different levels of abstraction can seem metaphysically inconsistent. For example, while a notion such as *pressure* can be reduced to the average behaviour of all molecules in a system, no single molecule can be said to possess, explain, or cause *pressure* in any meaningful sense. *Pressure* is simply a concept which exists at the macro-level, but not at the micro-level. Nor can *pressure* and *temperature* be decomposed separately (e.g., there are not two types of molecule which cause *pressure* and *temperature* independently), rather, the properties of the macro-level appear to *emerge*, fully formed, once the micro-level analysis becomes sufficiently complex. In more general terms, there is some point in our analysis at which the collection of molecules ceases to be, and is replaced by something radically different: a gas (see, e.g., Truesdell & Muncaster 1980).

Applying this analogy, if we allow that the relationship between the brain and phonology is one of *emergence*, rather than a strict compositional semantics, then a *Linking Hypothesis* should take the form of a complex dynamical system, and demonstrate the emergence of phonology-like properties from some specific combination of brain-like elements.

2. Introducing Attractors

The preceding argument leaves us with a well-defined problem: What kind of dynamical system could possibly give us something like a phonological grammar? The first obstacle to answering this question is that, while formal grammars are defined over a set of discrete symbols, dynamical systems (such as the brain) are typically understood as being fundamentally continuous³. This is where attractor dynamics are critical, because attractors allow us a way of realizing discrete behavior in an otherwise continuous system. Moreover, they are easily realizable in neural networks, making them a plausible candidate for a neural mechanism capable of underlying the discrete behaviour observable in phonological grammars.

² Conceptually at least, this analogy is not a novel idea in phonology. For example, it appears in Prince & Smolensky (1997) as a proposal for interpreting Optimality Theory.

³ In one sense, this situation is precisely the inverse of the kinetic theory of gasses, which seeks to replace many discrete particles with a continuous field (Truesdell & Muncaster 1980).

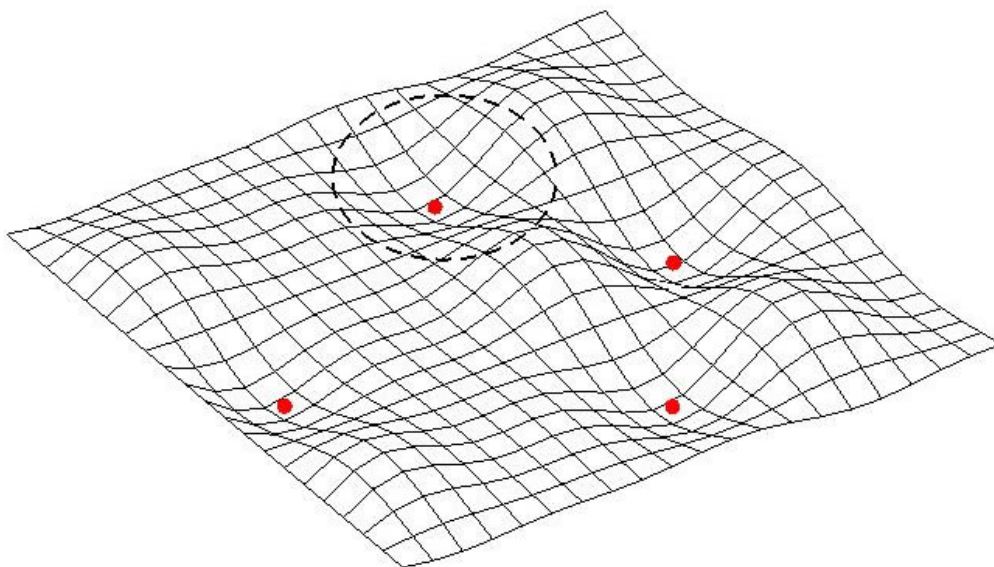


Figure 1: Conception of a network state-space. The z-axis corresponds to the free energy of the network. The red dots are attractors. (Illustration courtesy of Chris Eliasmith, Scholarpedia; source DOI: 10.4249/scholarpedia.1380; license CC BY-NC-SA 3.0.)

Like other artificial neural networks (ANNs), attractor networks consist of a number of simple units, which are interconnected with varying degrees of efficacy. Unlike other ANNs, attractor networks are characterized by symmetrical connections between units, which cause the network activity to settle on one of a number of asymptotically stable network states (i.e. attractor states). These stable states can be formally defined as local minima in an energy function and the behaviour of the network can be understood as analogous to the second law of thermodynamics: the entropy of the system increases over time, as the free energy decreases. This is sometimes visualised as a landscape of peaks and valleys (Figure 1), with the network always rolling down into the nearest valley.

The dynamics of attractor ANNs were popularized by Hopfield (1982), who noted that, if the attractor states are taken to represent pieces of information, then the network functions as a content addressable memory system.

Crucially for linguists, these attractor-memories are effectively discrete pieces of information. This is even true in cases where the individual units of the network are functionally gradient (Hopfield 1984). Thus, attractor dynamics are arguably our best candidate for explaining how a grammar over discrete elements could emerge in a seemingly analogue system like the human brain.

The model under examination here, the Phonological Latching Network (PLN), represents an attempted first step towards such a model. In its nascent form, it is necessarily an incomplete model of phonological grammar. It has no notion of lexical items, suprasegmental phenomena, or even a distinction between underlying and surface forms. Nonetheless, it does demonstrate how quintessentially phonological phenomena, such as assimilation, the Sonority Sequencing Principle (e.g., Clements 1990), and the Obligatory Contour Principle (e.g.,

McCarthy 1986), can emerge spontaneously from a relatively simple form of neural coding and memory retrieval.

3. Background and Outline of the Model

The PLN is a type of attractor network, similar to the Hopfield network (Hopfield 1982). This means that it stores memories as asymptotically stable states, which the network “self-organises” towards. However, most Hopfield-like ANNs have relatively simple dynamic properties: once switched on, the network will begin rearranging itself into the closest attractor state, where it will remain until the simulation is switched off. This limited degree of complexity has proven sufficient for investigating certain aspects of perception (e.g., Nasrabadi & Choo 1992) and memory capacity (e.g., Tsodyks & Feigelman 1988). However, it is clearly inadequate for modelling natural language grammar, which requires (minimally) the ability to define relationships between discrete memories.

Latching networks can be understood as an attempt to introduce between-memory dynamics into an attractor network. Fundamentally, latching networks behave like a Hopfield network, with the additional property that once an attractor state has been reached; the network begins to “latch” into a different attractor basin. Thus, the network can produce strings of discrete elements, which exhibit a kind of inherent grammar.

The latching dynamics themselves emerge from the introduction of a “fatigue” function (i.e. adaptation or inhibition) to active units, which means that attractor states become increasingly unstable once reached. This is what causes the network to latch into a different, nearby attractor, and ultimately places restrictions on what kinds of strings the network can produce.

3.1. *The Potts Unit*

The notion of fatigue in a latching network requires that individual units have an inactive state, that is a state which they tend to after periods of activity. The classical binary-unit Hopfield network lacks this property, since its units are either in an excitatory or inhibitory state.

The solution explored here is replace the binary-state Hopfield units with multi-state (or “Potts”) units, which have previously been studied as models of associative memory (Treves 2005; Russo & Treves 2012; Pirmoradian & Treves 2012; 2014; Song, Yao & Treves 2014; Kang et al. 2017; Naim et al. 2017). As in the case of the Hopfield network, single unit dynamics can be modelled using a rule based on heat bath dynamics (Treves 2005; Kanter 1988). These dynamics can be conceptualized as something akin to a compass needle being pulled in different directions by the various inputs received from other units in the network. The number of different directions in which the needle can be pulled is determined by the parameter S , which is typically in the order of 5 to 9, with one extra direction for the inactive state. Therefore, the state of a given Potts unit i is a probability vector of $S+1$ components, denoted below by σ_i^k for the active states, and σ_i^0 for the null-state.

At time t , the value for each active state k of any given unit i is given by the equation:

$$\sigma_i^k(t) = \frac{\exp[\beta r_i^k(t)]}{\sum_{l=1}^S \exp[\beta r_i^l(t)] + \exp[\beta(\theta_i^0(t) + U)]} \quad (1)$$

Where r is dynamic input variable, β is the global noise parameter, and U is a global parameter determining input to the inactive state. The time dependent thresholds for each state of each unit are given by the vector θ_i , which also has $S+1$ components denoted by θ_i^k for the active-state thresholds, and θ_i^0 for the null-state threshold.

Complimenting Equation 1, the value for the inactive state at time t is given by:

$$\sigma_i^0(t) = \frac{\exp[\beta(\theta_i^0(t) + U)]}{\sum_{l=1}^S \exp[\beta r_i^l(t)] + \exp[\beta(\theta_i^0(t) + U)]} \quad (2)$$

Calculating the values for σ_i at time t requires first determining both the values for the dynamic thresholds θ_i and the input variables r_i , which are linked through a system of differential equations (Equations 3, 4, and 5).

Firstly, the dynamic thresholds for the active-states are calculated from the current state of σ_i :

$$\tau_2 \frac{d\theta_i^k(t)}{dt} = \sigma_i^k(t) - \theta_i^k(t) \quad (3)$$

As the level of activation of a given state, k , in σ_i increases, so too will the corresponding threshold in θ_i , modulated by the coefficient τ_2 , which is a global parameter controlling the rate of active-state fatigue (or adaptation).

The dynamic threshold for the null-state is given by:

$$\tau_3 \frac{d\theta_i^0(t)}{dt} = \sum_{k=1}^S \sigma_i^k(t) - \theta_i^0(t) \quad (4)$$

Therefore, θ_i^0 increases relative to the sum of all active-states in σ_i , modulated by the global parameter τ_3 .

Note that θ_i^0 and θ_i^k (and their respective parameters τ_3 and τ_2) are intended to model two different forms of fatigue over two different timescales. While τ_2 is typically assumed to represent the rate of short-term depression in synapses, τ_3 is assumed to represent the rate of slow inhibition within a cortical patch.

Finally, once the dynamic thresholds for unit i at time t are known, the input variables r_i^k , can be calculated with respect to the local field h_i^k :

$$\tau_1 \frac{dr_i^k(t)}{dt} = h_i^k(t) - \theta_i^k(t) - r_i^k(t) \quad (5)$$

The local field for each state at time t is defined as the summed influence of presynaptic units, added to a local feedback term with the coefficient w :

$$h_i^k(t) = \sum_{i \neq j}^N \sum_{l=1}^S J_{ij}^{kl} \sigma_j^l(t) + w \left(\sigma_i^k(t) - \frac{1}{S} \sum_{l=1}^S \sigma_i^l(t) \right) \quad (6)$$

Where J_{ij}^{kl} denotes the connection strength between state k of unit i and state l of unit j (see Section 3.3.1 below for explanation of how connections strengths are determined).

Under the standard interpretation, each Potts unit is an effective model for a smaller attractor network (Naim *et al* 2018). Therefore, the w -term is intended to subsume the internal dynamics of each cortical patch. Continuing the compass needle analogy, it can be thought of as giving the compass needle an extra push towards whichever direction it is currently closest too.

3.2. Latching Dynamics

The relationship between fatigue on individual units and the emergence of latching dynamics is relatively transparent: an attractor state simply can't be maintained once the active units start switching off. What is less transparent however, is the rich complexity of the latching dynamics themselves.

In one sense, a latching network obeys the same principle of minimizing free-energy that all attractor networks obey, that is it "rolls into the valley" (Figure 1). The additional complexity arises from the fact the free-energy of any given network state is continuously changing as the fatigue rises and declines on individual units. In other words, the attractor landscape itself is constantly shifting. What was "downhill" at one moment in time can become "uphill" the next. The sheer mathematical complexity of these dynamics means that attempting to give a deterministic account of why one attractor latches into another is, although theoretically possible, massively intractable in practice.

For this reason, latching dynamics have more commonly been analysed probabilistically, for example, what determines the probability of a latch between any two attractors? This is still a non-trivial problem, but in general terms we can state that the probability of a latch between any two given attractors in the network depends on the overlap in the representations of those attractors (Russo & Treves 2012; Kang *et al.* 2017). The notion of "overlap" here has two dimensions: Firstly, how many active units do the two attractor states share? Secondly, how many of those shared units are in the same unit state? The interaction between these two types of overlap is quite complex, owing to the fact that they are governed by slightly different fatigue effects. The fatigue on individual unit states is controlled by the parameter τ_2 , while the fatigue on whole units is controlled by the parameter τ_3 . In the case where $\tau_2 \ll \tau_3$, an individual unit state will fatigue long before the unit itself begins to switch off (i.e. enter its inactive state). Thus, the degree of fatigue of an individual unit can bias the target of a latch in several ways: If a given unit is not fatigued, then the network will prefer to latch into an attractor in which that unit is both active and remains in the same unit state. However, if an individual unit state is fatigued, but not the whole unit, then the network might prefer to latch into an attractor in which the unit is active but in a

different state. Finally, if the unit itself is fatigued, then it will begin to enter to switch off and the network will prefer to latch into an attractor in which that unit is inactive.

The resulting global dynamics produces distinct “latching bands” in the degree of overlap between attractors: for some degrees of overlap, a latch will be highly probable, while for other it will be impossible (Russo & Treves 2012). If we allow ourselves a rhetorical simplification, we could say that the latching obeys a Goldilocks-principle; preferring to latch between memories which are neither too similar nor too dissimilar. In this sense a latching network always has an inherent grammar to it, since encoding multiple attractors in the network will always produce varying degrees of overlap between those attractors. Thus, a given latching network typically cannot produce all possible permutations of the memories represented by its attractors, but only a subset.

Finally, although the description of latching dynamics given so far only considers the probability of a latch between any two attractors, it should not be inferred that the network behaves like a finite-state machine. A latching network typically does exhibit long distance effects. This is a consequence of two facts: Firstly, the recovery time of a fatigued unit will typically be longer than a single latch. Thus, even if a given unit is inactive in the current attractor, it may still be fatigued from some earlier activation, and thus be less inclined to switch on again for the next latch. Secondly, in practice the retrieval of a memory is not actually understood as reaching one specific attractor state, but rather as passing through that state’s basin of attraction. This means that there are very many network states that would all be interpreted as a retrieval of the same memory, and each of these network states can behave differently in terms of where they would prefer to latch next.

When viewed from the macro-level then, the behaviour of the network might seem quite opaque: a single memory (or attractor basin) can produce a latch to one of many different targets, for reasons which are only apparent when viewed from the micro-level. This typically precludes reducing the global behaviour of the network to that of a deterministic automaton.⁴

Despite this, it is nonetheless possible to uncover distinct tendencies or biases in the strings produced by latching, when using probabilistic methods. As we shall see, the Goldilocks behaviour of the network can be seen to give rise to common phonological processes such as place assimilation and the Obligatory Contour Principle (OCP), while the slower cycles of fatigue can reproduce a kind of Sonority Sequencing Principle (SSP).

3.3. *Constructing a Neurologically Plausible Model*

Unlike many ANNs, the Potts units of the latching network do not strive to model individual synapses, firing rates or action potentials. Rather they can be thought of as an effective, or “grey box”, model, where certain details are subsumed into a system of differential equations. For this reason, a Potts model is as much a

⁴ This does not entail that *no* configuration of a latching network can reproduce *some* level of complexity on the Chomsky hierarchy; this ultimately remains to be seen.

theoretical model of specific system dynamics, as it is a model of neurological reality. Indeed, while many aspects of the latching model are intended to capture known facts about neural function, the exact neural implementation of a Potts unit is somewhat open to interpretation.⁵ Under the standard view, each Potts unit is an effective model for small patches of cortex. The active states of each unit represent different local attractors in each patch, while the self-reinforcement term represents the internal attractor dynamics of the patch. Then the behaviour of the network as a whole is taken to model global dynamics between relatively distant areas of the cortex (Naim et al. 2018). This standard view of a Potts network seems well suited to modelling language, which is known to be a widely distributed cognitive faculty (see, e.g., Hickok & Poeppel 2007).

The PLN is intended to model the representation of phonological information in the cortex. While a great deal is still unknown on this topic, recent ECoG studies have uncovered a striking degree of congruence between phonological representations in the cortex and the abstract, discrete features employed by linguists to explain the behaviour of phonological grammars (Bouchard et al. 2013; Mesgarani et al. 2014). Specifically, these studies uncovered the existence of small patches of cortex which are highly sensitive to specific phonological features. Moreover, they hint at a spatial asymmetry between manner and place features, with manner features being distinguished more strongly in the superior temporal gyrus (STG), and place features being distinguished more strongly in the ventral sensorimotor cortex (vSMC). Similarly, both experimental results and theoretical modelling have suggested that phase coupling between these areas may form a critical component of the phonological capacity (Assaneo & Poeppel 2018).

These findings suggest three relevant criteria for the structure of the PLN: Firstly, the network should be split into two sub-networks: an auditory sub-network for manner features, and a motor sub-network for place features, and that production should arise from synchronous activity between these areas. Secondly, phonological similarity between phones should be captured in terms of shared units in the network (i.e. shared patches of cortex), such that the Goldilocks-principle is acting over phonological properties. Finally, the congruity between the ECoG studies and phonological theory suggests that the representations themselves could be constructed using abstract phonological features as a guide.

3.4. *Building Phones*

Unlike neural networks typically employed in machine learning and connectionist frameworks, the PLN is not subject to any form of supervised learning (cf. Alderete & Tupper 2018). Rather, the patterns of activity which represent memories are generated algorithmically by the experimenter, and then encoded in the connections between units using a simple Hebb-like rule.

⁵ In Marrian terms (e.g., Marr & Poggio 1973), if the linguistic model is the *computational* level, then the latching network is the *algorithmic* level, while the *implementational* level would be occupied by some exact neural model of the Potts units.

Because the memories in the PLN are intended to represent phones, the algorithm for memory generation in the PLN works from a given phoneme inventory, which is formally defined in terms of a relevant set of phonological features (see Appendix). Broadly, each of the features is defined as a random pattern of activity. These patterns can then be combined into phones, following the definitions in the phoneme inventory. The process for combining features is a competitive one, whereby the individual features are used as competing “suggestions” for the final phone. Contradictions between suggestions are resolved by weighting individual features, such that only the strongest suggestions for each unit will contribute to the phone representation.

The same features are used in both the auditory and manner sub-networks, and the asymmetry is achieved by reversing the weighting of those features. So, the auditory network representations are generated with heavily weighted manner features and weakly weighted place features, and vice-versa for the motor sub-network.

The phone inventory is loosely derived from English phonology, with the important caveat that there are no minimal pairs based on voicing distinction. The large number of features means that phones are redundantly over-specified, as otherwise the algorithm tended to produce phones with excessive overlap. Slowly adding redundant features to the inventory was a way of overcoming this problem. However, it should be noted that some information is lost during phone creation, so not all the features should be regarded as playing a role in the behaviour of the system (by extension, the PLN should not be interpreted as for or against any particular theory of phonological features).

The process for generating representations in the PLN will now be described in detail. First, each phone μ is formally defined as a set of M features:

$$\mu := \{\varphi_1^\mu, \varphi_2^\mu, \dots, \varphi_M^\mu\} \quad (7)$$

The notation φ^μ indicates that feature φ is a member of phone μ .

The features defining a given phone are, in principle, unordered. However, the process for generating phones requires two different orderings of the features in μ (one for each sub-network).

A sub-network is defined as a pool of units and is denoted by Q , which in the PLN can take the value *mot* or *aud*. Any given unit in the network, i , is assigned membership to one, and only one, of the pools. The two pools contain the same number of units: $N/2$.

The auditory and motor components of each phone are defined as ordered tuples of all elements in μ :

$$\mu^Q := \varphi_1^{\mu^Q}, \dots, \varphi_m^{\mu^Q}, \dots, \varphi_M^{\mu^Q} \quad (8)$$

The order is always derived from the inventory given in the Appendix. Also note that μ^{aud} and μ^{mot} always contain the same elements, but in the reverse order, that is the relationship always holds that $\varphi_m^{\mu^{mot}} = \varphi_{M-m+1}^{\mu^{aud}}$.

The function W assigns a weight to each feature, with respect to its position in μ^Q , such that:

$$W(\varphi^{\mu^Q}) = e^{\frac{q(m-1)}{M-1}} \quad (9)$$

Where m is the index of feature φ in μ^Q , M is the total number of features in μ^Q , and q is a global parameter used to control the cumulative influence of lowly weighted features (the smaller the value of q , the greater the influence of the lower weighted features).

The result of the function W is that the weightings of the features in μ^Q fall along an exponential scale between 1 (when $m=1$) and e^q (when $m=M$).

The weightings from W are used to determine the actual representations for a phone.

First, the representation for phone μ in pool Q is denoted as ξ^{μ^Q} , which is defined as a tuple whose components represent the units in pool Q , and can take a value from 0 to S .

$$\xi^{\mu^Q} := \xi_1^{\mu^Q}, \dots, \xi_i^{\mu^Q}, \dots, \xi_{\frac{N}{2}}^{\mu^Q} \quad (10)$$

The final representation for a given phone will simply be the concatenation of the two pools: $\xi^\mu := (\xi^{\mu^{mot}}, \xi^{\mu^{aud}})$.

Generating the representations for phones depends on the representations for individual features. Each of the features in the phoneme inventory is defined as a hypothetical network state within each sub-network which, following Pirmoradian & Treves (2012; 2014), are generated using sparse⁶ patterns of noise. The random noise pattern representing a feature φ is indicated as ξ^φ , where, again, each element takes a value between 0 and S .

$$\xi^\varphi := \xi_1^\varphi, \dots, \xi_i^\varphi, \dots, \xi_{\frac{N}{2}}^\varphi \quad (11)$$

Crucially, the patterns for features are uncorrelated with one another, i.e. they should be approximately equally dissimilar.

Additionally, the sparsity of these patterns is enforced by the parameter a_{feat} , which represents the probability that the value of any component ξ_i^φ is non-zero. In practice, the value of a_{feat} is typically lower than the value of a , the sparsity of the phones. This ensures that no phone can be dominated by a single feature.

Note that any given feature pattern, ξ^φ , is constant for all phones and all pools. Features vary only in terms of their membership in μ and weighting in μ^Q . Also note that each feature pattern is only defined over half the total units of the network. This is because, in principle, each feature appears in both the auditory and motor sub-networks.

⁶ That is, only a small subset of units are active.

As well as the patterns representing phonological features, each phone also has a corresponding “noise” feature, \mathcal{N} , which is introduced as a means of preventing excessive overlap between phones. The noise feature is similarly defined:

$$\xi^{\mathcal{N}} := \xi_1^{\mathcal{N}}, \dots, \xi_i^{\mathcal{N}}, \dots, \xi_{N/2}^{\mathcal{N}} \quad (12)$$

Having defined and generated all the relevant feature representations, the final value of any unit in ξ^{μ^Q} is set to the value of k (between 0 and S) which carries the highest weight, from W , which is summed over all features in phone μ .

$$\xi_i^{\mu^Q} = \arg \max_{1 \leq k \leq S} \sum_{\varphi \in \mu} \delta_{\xi_i^{\varphi} k} W(\varphi^{\mu^Q}) + p e^q \delta_{\xi_i^{\mathcal{N}} k} \quad (13)$$

The Kronecker delta is a function which equals 1 when its arguments are the same, but 0 otherwise. The last term in (13) represents the influence of each phone’s unique noise feature, \mathcal{N} , where p is a global parameter used to control the influence of all noise features. Note that if $p=1$, then the weight of the noise feature will be equal to the weight of the strongest feature in μ^Q . High values of p (greater than 1), were found to be useful for maintaining an optimum degree of overlap between representations.

Additionally, the sparsity of the representations is maintained by assigning a value of 0 to those units whose weighted suggestion falls below some threshold. The value of this threshold depends on the sparsity parameter a , such that only the $aN/2$ strongest suggestions in ξ^{μ^Q} are non-zero.

Having generated the representations for each phone, the patterns are encoded in the weight matrix as attractors using a Hebb-rule. Each phone μ suggests a connection strength J between state k of unit i and state l of unit j , which is given by the rule in:

$$J_{ij}^{kl}(\mu) = (\delta_{\xi_i^{\mu} k} - \frac{a}{S})(\delta_{\xi_j^{\mu} l} - \frac{a}{S})(1 - \delta_{k0})(1 - \delta_{l0}) \quad (14)$$

Here, as before, the Kronecker delta’s output is 1 when the two arguments are equal and is 0 otherwise. Therefore, in a pattern, ξ^{μ} , if unit i is in state k and unit j is in state l , where $k=l$, then the connection will be positive (excitatory), else the connection will be negative (inhibitory). The last two factors ensure there are no connections to/from units in the null state (if k or l equal 0).

The final value for each connection is determined by summing over all memories in the network, and multiplying by a normalization factor:

$$J_{ij}^{kl} = \frac{c_{ij}}{Ca(1-\frac{a}{S})} \sum_{\forall \mu} J_{ij}^{kl}(\mu) \quad (15)$$

Where c_{ij} is set to 1 when i and j share a connection and is 0 otherwise. This value is normalized by C , the average number of connections per unit, and a , the sparsity parameter.

The probability that they share a connection is defined by the variable c_{int} if i and j are both in the same sub-network, or c_{ext} if they are not:

$$\text{For } Q \neq R, \quad c_{ij}^{QR} = \begin{cases} 1 & \text{with probability } c_{ext} \\ 0 & \text{with probability } (1 - c_{ext}) \end{cases}$$

$$\text{For } Q = R, \quad c_{ij}^{QR} = \begin{cases} 1 & \text{with probability } c_{int} \\ 0 & \text{with probability } (1 - c_{int}) \end{cases}$$

This process is intended to ensure that the similarity between the representations of phones in the PLN correlates strongly with their phonological similarity, as is given by the feature definitions in the phoneme inventory. We can see evidence of the non-random structure of the PLN memories, shown in *Figure 2*. Here we can see that, in general, the more units two memories in the PLN share, the more likely it is that those shared units are in the same Potts state. This implies that overlap between representations is a consequence of shared features which suggest specific Potts states for individual units.

4. Analysis of PLN Behaviour

Because the process of generating features depends heavily on randomization, it is possible to generate multiple weight matrices for the same phoneme inventory, which have different latching properties (i.e. they produce different grammars).

Using the same phoneme inventory and network hyperparameters (see Appendix), the latching strings from 125 trials, representing 8 different grammars,

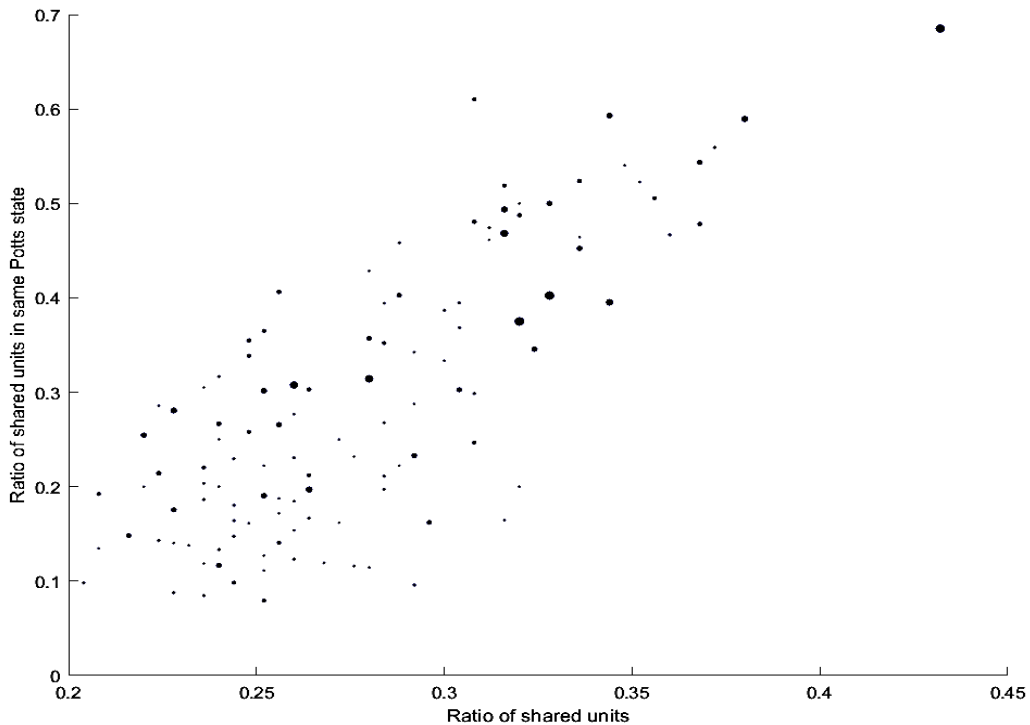


Figure 2: Overlap of memories produced by feature super-position. The size of each circle indicates the total number of attested transitions between the two memories during the simulations.

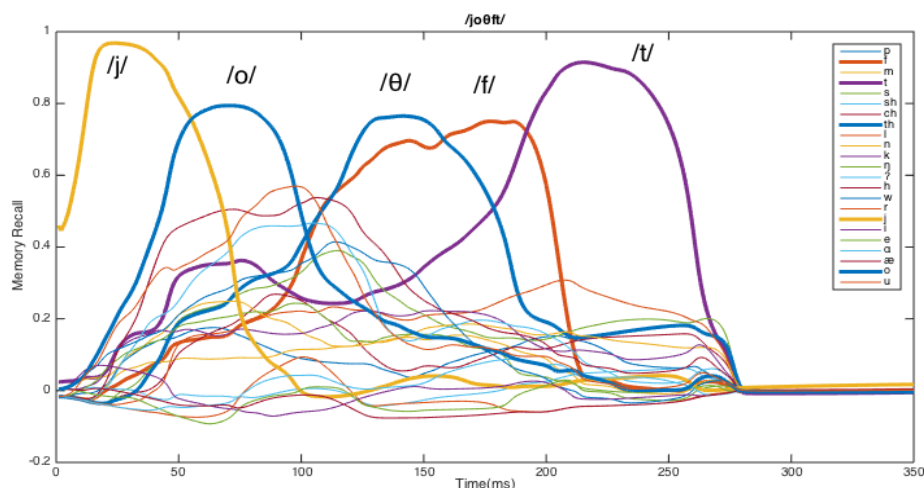


Figure 3: Example of a latching string.

were collected into a corpus containing a total of 464 individual phoneme transitions. This was found to be large enough to allow statistical generalization, but small enough that all latching transitions could be manually checked for network pathologies (failed retrievals, mixed states, etc). Only strings which exhibited no obvious pathologies were included in the corpus. All strings were between 2 and 8 segments long, with an average length of 4.7 segments. Strings were generated by placing the network into a state which matched a 50% memory retrieval and allowing it to run for 400 time steps.

The strings were assessed for evidence of assimilation, OCP and SSP. The rate at which these phenomena occur was then compared to chance level, that is a grammar in which the probability of a transition between any two phones is the same for all phones in the inventory. The extent to which the PLN grammars deviate from chance level can be taken as evidence of whether these processes are inherent to the PLN.

4.1. Segmental OCP

In its general form, the Obligatory Contour Principle (OCP) requires that there be some minimum degree of difference between adjacent objects. This may or may not be an instance of a more general bias against repetition in language (see, e.g., Walter 2007). In relation to segmental phonology, this can be interpreted in two different ways: firstly, it can mean that the same phone cannot surface twice in a row, or secondly, that adjacent segments cannot be similar with regards to some featural specification (McCarthy 1986).

This first sense of segmental-OCP is a trivial property of the PLN, since the latching dynamics are driven specifically by an active memory becoming unstable. There is simply no way the network could latch out of, and immediately back into, the same memory. The simulations confirmed this, with phone repetitions exhibited in exactly 0% of the recorded transitions.

The PLN also seems to exhibit something closer to the second definition of segmental-OCP. For example, there were no recorded examples of a transition

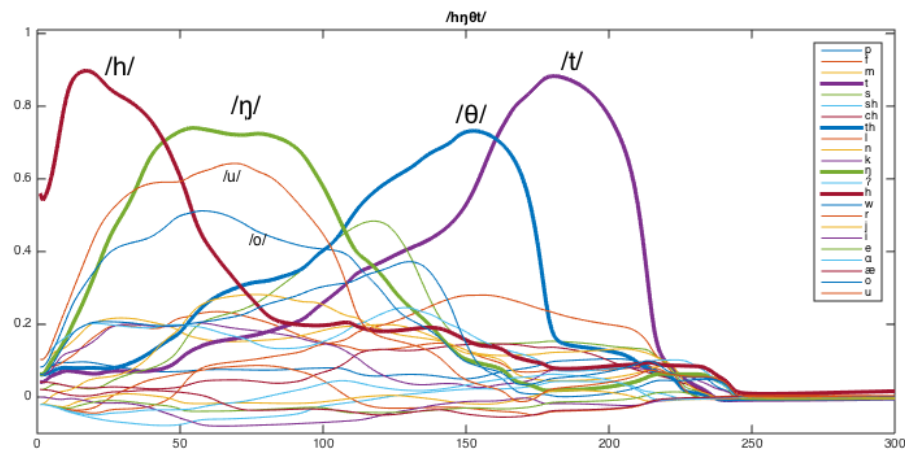


Figure 4: The /θ/ and /t/ phones are similar in both their manner and place of articulation, but are still a possible transition for the PLN.

between /s/ and /ʃ/, suggesting that the network has reproduced something like the OCP-driven epenthesis seen in English plurals and possessives (e.g., bu[ʃ] → bu[ʃəz] etc.). However, one grammar did spontaneously produce the string [kɪtʃθu], where the transition from /tʃ/ to /s/ would normally be seen as an OCP violation in the context of English phonology.

A closer examination of the representation overlap of these phones reveals the important difference. Firstly, the total percentage of shared units between /s/ and /ʃ/ in this grammar is much higher (31.2%) than /s/ and /tʃ/ (22.4%). And secondly, of those shared units, a much higher percentage are in the same Potts state when comparing /s/ to /ʃ/ (50%) than /s/ and /tʃ/ (28%). This supports the hypothesis the absence of /s/ → /ʃ/ transitions in the PLN is an OCP effect, while /s/ and /tʃ/ are dissimilar enough to fall within the “Goldilocks” zone.

4.2. Assimilation

Processes in which segments become more similar to their neighbours – in terms of their feature specification – are extremely common cross linguistically (e.g.,

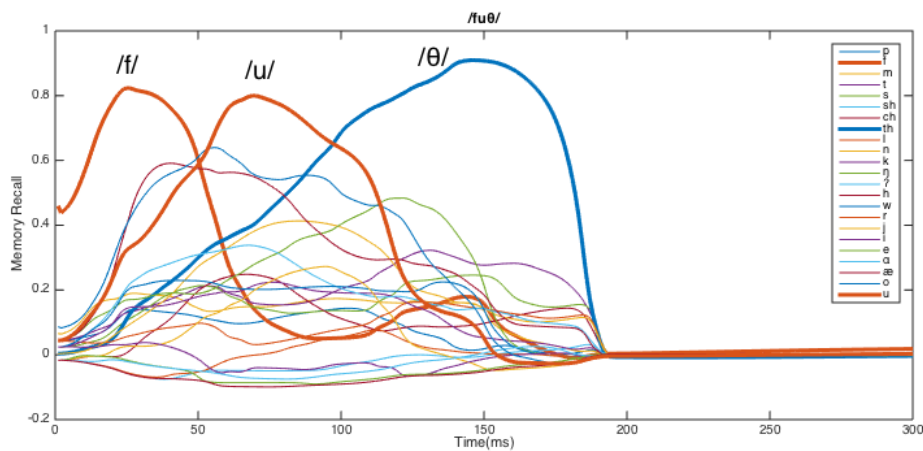


Figure 5: The /f/ and /u/ share the feature [round], so the first transition is interpreted as an instance of place assimilation.

Ohala 1990). With the PLN, a transition was counted as an instance of assimilation if the two phones shared a feature, as defined by the inventory in the Appendix. An example of this is shown in *Figure 5*.

4.2.1. Place

Transitions exhibiting place assimilation were found in 244 (52.6%) transitions, which is slightly above the chance rate (44%). However, the picture becomes more interesting when we break down the assimilation probabilities for each feature. As we can see in *Table 1*, the features HIGH, EXTERIOR, LABIAL, VELAR and ALVEOLAR appear to assimilate at above chance rate, while the others assimilate below chance rate.

These numbers suggest that only some of the features are participating in assimilation. This is arguably a welcome result, since natural phonological grammars typically only exhibit assimilation for one or, at most, a few place features.

However, these numbers alone do not immediately inform us of why some features participate in assimilation, but not others. This picture is further complicated by the fact that not all of these features are independent. In cases where the phones delineated by one feature are a strict subset of the phones delineated by another feature (e.g., all labials are also exterior, etc.), then a naive statistical method doesn't allow us to determine which feature is decisive for causing assimilation.

We can partially circumvent this problem by comparing mutually exclusive pairs of features, for example, HIGH vs LOW, LABIAL vs CORONAL, and ALVEOLAR vs POST-ALVEOLAR. Each phone may have, at most, one of the features from each of these pairs.

Looking at *Table 1*, we can see that within each exclusive pair, it is the feature with the highest weight during phone generation (Section 3.3.1) which appears to assimilate at above chance rate, while the feature with the lower weight assimilates at a below chance weight.

FEATURE	ASSIMILATION %	BASELINE %
HIGH	3.66	1.7
LOW	1.08	4.73
FRONT	3.45	6.8
EXTERIOR	35.34	18.9
LABIAL	15.73	6.8
DORSAL	2.16	6.8
CORONAL	9.27	12.1
VELAR	2.59	0.76
GLOTTAL	0.22	0.76
ANTERIOR	0.65	1.7
ALVEOLAR	6.03	4.73
POST-ALVEOLAR	0.43	0.76

Table 1: Place assimilation probabilities by feature, ordered from strongest weight in motor sub-network (HIGH) to lowest (POST-ALVEOLAR).

FEATURE	ASSIMILATION %	BASELINE %
APPROXIMANT	0.22	2.01
CONTINUANT	76.29	54.63
NASAL	3.01	1.7
SONORANT	61.42	31.94
VOCALIC	17.89	7.05
CONSONANTAL	26.5	37.05

Table 2: Manner assimilation probabilities by feature.

This gives us some indication that the relative weighting of features during phone creation plays a role in determining assimilation in the emergent grammar. Intuitively, this makes sense insofar as features with heavier weights will “suggest” more unit states for the final representation of each phone. Therefore, the heavier the weight of a feature, the more overlap we should expect between any two phones which share that feature, and the greater the probability that the network will prefer to latch between them.

4.2.2. Manner

The random baseline for manner assimilation is much higher at 81.1%, owing to the smaller number of manner features, and the larger number of individual phones delimited by each manner feature. The actual rate of manner assimilation within the network is, again, slightly above chance at 89.4%.

Similar to place features, we also see a difference between individual manner features (Table 2).

That the CONTINUANT and NASAL features exhibit assimilation is broadly in keeping with natural phonology, for example, intervocalic spirantization (Kaplan 2010) and vowel nasalization (Krämer 2019). More surprising, perhaps, is the apparent assimilation of the features SONORANT and VOCALIC, which are typically not thought to spread or assimilate (see, e.g., Clements & Hume 1995 where these features appear on the root node). However, this can actually be explained as an effect of the sonority sequencing effect in the network (see Section 4.3), whereby the network tends to slowly oscillate between greater and lesser sonority. Since the features SONORANT and VOCALIC are the main delineators between degrees of sonority, the sonority sequencing will naturally cause phones with these features to cluster together, rather than being even distributed. Thus, the statistical effect need not be regarded as a consequence of spreading or assimilation *per se*, but rather of sonority sequencing.

4.3. Sonority Sequencing Principle

The Sonority Sequencing Principle (SSP) refers to the tendency for sonority to follow a monotonically rising-then-falling pattern across a single syllable. Arguably, this forms the very definition of a syllable: it is a sonority peak (Clements 1990). For this reason, the SSP represents a good measure for the “naturalness” of the strings produced by the PLN. For example, strings which

VOWELS GLIDES LIQUIDS NASALS OBSTRUENTS

0	1	2	3	4
---	---	---	---	---

Table 3: Sonority scale.

neatly transition between consonants and vowels could be regarded as more natural than strings which consist only of stops.

Unlike the other measures, the extent to which the network obeys sonority sequencing is defined in relation to whole syllables, not individual transitions. And since the PLN does not itself process any information relating to syllable structure, the experimenter must parse the strings into syllables manually. This requirement presents the basis for a simple metric for approximating the model's preference for strings which obey SSP. Specifically, each string produced by the PLN is given the best possible parse according to the SSP. The string is then assigned a value from the sonority scale (Table 3), according to the *least* sonorant nucleus required when parsing (Table 4).

Note that this method ignores syllable plateaus and size of the sonority “jump” between adjacent segments. Some examples of how these scores would be assigned to example strings are given in Table 4.

Once every string in the database has been assigned a sonority score, the mean score (across all strings) is compared to a random baseline, whose sonority sequencing score has been computed for strings of length 3, 4, 5, 6, 7.⁷ The sonority scores for different string lengths, both from the PLN and the baseline, are given in Figure 6.

The SonSeq score for the latching strings is lower than the baseline for all string lengths, suggesting that the PLN tends towards strings which can be parsed by to the SSP.

Naturally, this simple metric inherently ignores various complexities associated with sonority sequencing in natural grammars (minimum/maximum distance, permissible plateaus, onset/codas asymmetries, etc.). However, it does capture the extent to which the PLN wants to oscillate monotonically between vowels and obstruents. This is informative insofar as it presents an unbiased measure of how well the latching strings conform to sonority sequencing, within the confines of a system which has no actual notion of syllable structure.

STRING	SYLLABLE PARSE	LEAST SON. NUC.	SONORITY SCORE
“Σ L O”	fl̩	o	0
“L Σ O”	l̩.fo	l	2
“Θ N Æ L P F”	θnælpf	æ	0
“Θ N Æ L P F M”	θnælp̩fm	m	3

Table 4: Example sonority scores.

⁷ Note that the SonSeq score worsens (increases) as the strings lengthen by simple virtue of the fact that the longer the string, the greater the probability of encountering a low sonority nucleus.

Mean SonSeq Score by String Length

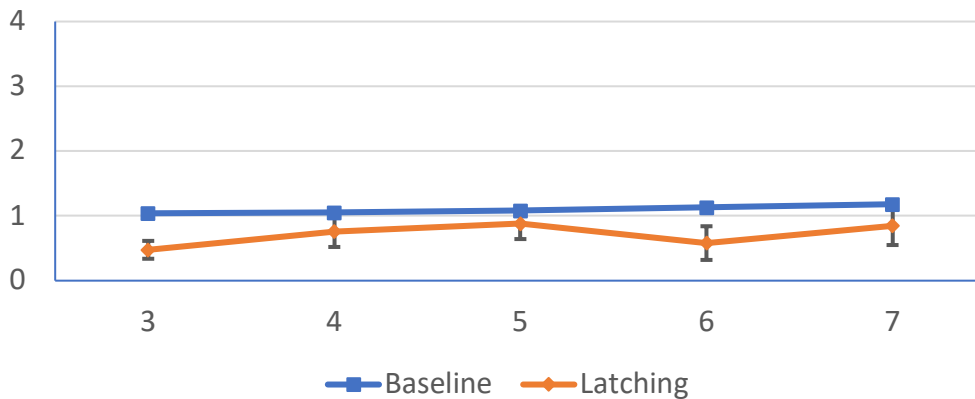


Figure 6: Sonority Sequencing score for latching strings (red) versus random baseline (blue).

4.3.1. SSP as Oscillation

Having established the PLN’s propensity for oscillating between sonorous and non-sonorous segments, it remains to determine *why* the network exhibits this behaviour. Much like the OCP effect, the SSP effect can also be understood as following from the fatiguing of individual units. In simple terms, because the network representations are intended to reflect phonological properties, we should expect that certain units will be more active when representing sonorous phones than non-sonorous ones (and vice-versa). Thus, if these “sonority units” are fatigued from repeated activation, then we should expect the network to latch into non-sonorous memories for a time, at least until the “sonority units” have recovered from their fatigue. Similarly, the converse will be true for any “non-sonority units” which are most active for non-sonorous phones. Therefore, we should expect the network to slowly oscillate between sonorous and non-sonorous states, driven by the slow fatiguing and recovery of the individual units.

Of course, this oscillation can only persist if sonority is indeed encoded in the network in this way. As already noted in Section 4.2.2, the degrees of sonority within the phone inventory are determined primarily by the features SONORANT and VOCALIC. However, because the process of generating representations relies on randomisation, we need to look at the network representations themselves to see whether or not these features actually play a role in producing the SPP effect. We can get a sense of this by grouping the individual phones in the network into 3 broad sonority categories: vowels, sonorant consonants, and obstruents (which correspond to the features SONORANT+VOCALIC, just SONORANT, or neither, respectively) and examining the average representation overlap within and across these categories.⁸

The data in Table 5 show the overlaps across these categories from a single randomly chosen grammar of the PLN. As we might expect, the average overlap

⁸ Distinguishing the entire sonority hierarchy requires additionally the features APPROXIMANT and NASAL. However, for legibility we can restrict ourselves to this tripartite distinction.

SONORITY CATEGORIES	AVERAGE % OF SHARED UNITS	% OF THOSE SHARED UNITS IN SAME STATE (ABSOLUTE %)
SON-SON	29.24	30 (8.8%)
OBS-OBS	27.6	31 (8.55%)
V-V	27.6	33 (9.2%)
SON-V	26.05	25 (6.47%)
OBS-SON	25.63	25 (6.38%)
OBS-V	24.83	20 (5.05%)

Table 5: Overlap across sonority categories within a single grammar.

is highest within each category (obstruent, sonorant, vowel), somewhat lower when comparing obstruents to sonorants and sonorants to vowels, and lowest when comparing obstruents to vowels. The divide is even sharper when we examine the ratio of those shared units which are in the same Potts state, where we also see a much higher ratio of shared unit states within categories, when comparing across categories (Table 5).

This pattern, taken with the high rate of SONORANT and VOCALIC assimilation (Section 4.2.2), supports the oscillation explanation outlined above. To understand why, recall that the network has two types of fatigue, one which applies to individual Potts states, and one which applies to whole units. The tension between these two types of fatigue are critical for determining the behaviour of the latching network. Specifically, latching is driven by memory overlap in the case where unit fatigue is slower than individual Potts state fatigue (Kang et al. 2017), which is the case in the PLN. This is because latching occurs when an attractor becomes unstable due to fatigue, and since unit states fatigue faster than whole units, then latching will be driven the competing drives to maintain active units but to deactivate fatigued unit states. The consequence in this case will be a latch between memories which share the most units, but only if those units differ enough in their individual states.

5. Discussion

The analysis of the latching corpus presented here suggests that the PLN exhibits a degree of place assimilation and sonority sequencing, with a near-absolute kind of segmental OCP, or anti-adjacent-repetition of phones.

In terms of understanding why the network exhibits certain behaviours, arguably the most straightforward of the three is the segmental OCP. The “Goldilocks” behaviour of the PLN—preferring latching targets which are sufficiently dissimilar but not too dissimilar—will naturally prohibit latching out of and back into the same phone. Of course, depending on the specific overlaps of the memories in the network, this OCP effect can also to extend to phones which are similar though not identical. Thus, as seen in Section 4.1, it is perfectly possible to create an English-like grammar where /s/ and /ʃ/ are separate phones, but where transitioning from one to the other is strictly impossible, by virtue of the high degree of overlap in their representations.

Similarly, the PLN's bias towards assimilation can be straightforwardly understood as a result of the "Goldilocks" principle – the network prefers latching targets which are sufficiently different from the current state (OCP), but not *too* different (assimilation). Once again, whether or not a given grammar actually exhibits a given type of assimilation depends on the exact network representations that constitute the phones in the inventory: If two phones share a feature with a higher weight (during phone creation), then more of the overlap between the phones will be determined by that feature, ergo strongly weighted features are more likely to cause assimilation.

Finally, the PLN's apparent preference for oscillating between greater and lesser sonority can also be understood as a cumulative effect of the fatiguing of individual units in the network. However, unlike the OCP and assimilation effects, we need to consider the role of fatigue over a longer timescale.

Nonetheless, because the PLN is, in some sense, an incomplete model of phonological processing, a certain degree of care is required when attempting to draw direct comparisons with concepts taken from phonological theory. With that in mind, it is worth considering some of the limitations of the PLN model, how that affects our interpretation in phonological terms, and what that might mean for future research.

For example, the OCP-like effect exhibited by the PLN does not, by itself, capture the variety of different phonological effects which phonologists might ascribe to the OCP. This is true even if we ignore suprasegmental phenomena (tone, etc.) of which the PLN has no notion. Indeed, even at the segmental level, we might cite the OCP as a motivator for epenthesis, deletion, gemination, metathesis, etc. But whether or not the PLN can exhibit any of these processes is a moot point, since they are defined as the relationship between a surface form and a corresponding underlying form, whereas the PLN has only a single level of representation.

However, this should not be regarded as a fatal flaw in the PLN *per se*, but rather as an indication of how the PLN should be expected to interact with the other components of a complete linguistic system. Speculatively, if the representations in the PLN were interpreted as surface phonological representations, then the underlying representations should correspond to the lexical representations which trigger a given latching string. In this way, input-output mappings in the phonology could be understood as the interaction between the lexical input and the PLN itself.

Again, the PLN does not have a lexical-memory component, so exactly how the activation of a lexical item triggers a latching string is not yet modelled explicitly. But the possibilities here are clearly bounded. For example, the PLN simulations are conducted by "giving" the network a single, incomplete pattern. The exact properties of this initial pattern are what determine the trajectory of the subsequent string. Moreover, it has already been established that small differences in the initial pattern can produce large differences much later in the string—an effect loosely analogous to a butterfly's flapping wing causing a hurricane on the other side of the world. For example, consider these three strings, taken from the same grammar in the PLN corpus:

- (1) a. ? m u o i f n m
 b. ? m u o i s n m
 c. ? m u o a f n m

Each string begins with an incomplete version of the same phone, /?/, and the strings follow the same trajectory for the subsequent 3 latches, before diverging at the 4th and 5th latches, and then returning to the same trajectory for the final two latches. Note that the cause for the differences in each string lies solely in the subtle differences in the initial state for each case, which are invisible when the system is viewed from the macro-level (recall: memory retrieval is understood as passing through an attractor basin, not arriving at an exact point).

This presents an obvious hypothesis that lexical items could trigger a given string simply by sending a short, initial cue to the phonological system. If we suppose that one such cue is sent every time, for example, the syntax/morphology picks a new morpheme, then the cues sent to the phonology would correspond to word/morpheme boundaries, and phonological processes could be understood as the latching network resolving the mismatch between the input from syntax/morphology and its own internal bias for preferred latching targets.

To give an explicit example, suppose we have a network which has latched into an /f/, and then receives a new initial cue in the form of a /z/, as in the case of an English plural like bu/f-z/. If, in the given language, the representation for these two phones are too similar, then directly latching into the /z/ will be impossible. Therefore, the network could react in a number of ways. For example, additional excitation might lengthen the duration of the current retrieved memory (gemination), the network might latch to a similar but sufficiently different memory (dissimilation), it might latch to an intermediate memory before latching to the /s/ (epenthesis), or might fail to latch to the /s/ entirely (deletion). Exactly which strategy the network adopts will depend on the exact nature of the input received from the lexicon. Thus, the phonological grammar for a given language would be localized both within the PLN, and the connections to the lexicon themselves.⁹

Whether or not this model is workable in practice is a topic for future research, since it presupposes a model of lexical storage and retrieval. Currently, there exists no method for exactly “controlling” the strings produced by a latching network. In part, this is because the number of possible initial states for the network is unfathomably large, 8²⁰⁰ in the case of the PLN (which is a number 180 digits-long if expressed in regular notation). However, while it is quite conceivable that the majority of those possible initial states do nothing interesting, it need only be true that a tiny subset of them produce unique strings in order for the PLN to be able to produce a vocabulary of lexical items which is comparable in size to that of a typical adult speaker (i.e. in the order of 10s of thousands).

Finally, it should be noted that the method for producing representations, outlined in Section 3.3.1 is somewhat volatile, insofar as it frequently produces grammars with obvious pathologies (failing to retrieve phones, mixed-state

⁹ Conceptually, this is strongly analogous to the Optimality Theoretic concepts of markedness (PLN representation) and faithfulness (connection to lexicon).

retrievals, etc.). The solution pursued here was to produce large numbers of grammars and filter out the pathological cases before conducting the analysis. However, in addition to being time-consuming, this method does not allow for a detailed analysis of exactly which variables distinguish the pathological cases from the phonology-like cases. A preferred approach would be the development of a memory-generating algorithm which allows for a more exact control over the variables that differentiate the possible configurations of the network. Such an algorithm has been developed in the context of semantic memories (Boboeva et al. 2018) but has not yet been generalised to a phonology-like case. Of course, semantic memories are fundamentally different to phonological memories insofar as the semantic system is much larger and depends on radically different associations between those memories. However, it is quite conceivable that the method employed by Boboeva et al might be modified for a smaller phonology-like system. This remains a plausible topic for future research.

6. Conclusion

At the start of this paper I claimed that the PLN can be understood as a *Linking Hypothesis* which bridges the ontological incommensurability between neuroscience and phonological theory. It does not do so by decomposing specific linguistic models into simpler computational mechanisms, but rather by demonstrating how to produce strings which exhibit phonology-like behaviour (assimilation, OCP, SSP), using only a small number of brain-like ingredients (recurrent connections, distributed representations, short-term adaptation), plus a system of memories defined in terms of phonological features. In this way, the components of the linguistic formalism are understood to be emergent from a complex dynamical system.

The relevance of the results from the model can be understood from two perspectives: that of the neuroscientist and that of the linguist. From the neuroscientist's perspective, it is significant that the phonological behaviours exhibited are not explicitly taught to the network, nor are they pre-programmed in any way. Rather, they seem to emerge spontaneously from the specific combination of phonologically-inspired representations and neurally-inspired network dynamics. This fact supports the plausibility of latching dynamics as a real neural mechanism. This type of indirect evidence is crucial because, although latching dynamics have been studied theoretically in a variety of contexts, measuring them directly is likely beyond current neuroscientific techniques. Of course, the PLN still leaves open a number of questions about the underlying neurological reality. Most notable is the specific neural correlate of the Potts units themselves, which are intended to subsume a large amount of potential complexity into a relatively simple and tractable approximation. However, the Potts units are not totally opaque, and the specific parameters of the model implicitly delimit the range of possible underlying biological mechanisms that we can posit. Further research into the PLN is likely to yield clearer predictions in this regard, because as the parameters of the model become more fine-tuned, so too do the neural predictions. Thus, the PLN presents us with an interesting case

where linguistic facts could be used to deduce relatively fine-grained neural properties.

From the linguist's perspective the implications of the PLN are less direct, since we are discussing across two quite different levels of abstraction, that is linguistics and neuroscience. In general, we should be cautious about drawing direct correlations between the ontologies of neutrally inspired models and formal linguistic theories. However, the PLN could nonetheless inform the discussions and assumptions *surrounding* formal linguistic theories, if not the theories themselves. One example of this is the topic of innateness and learnability which, although not necessarily properties captured within a formal theory, are nonetheless topics of thorough debate by linguists (e.g., Odden 2013).

Indeed, under one reading, Chomsky's articulation of Universal Grammar (UG) could lead one to believe that the primary goal of formal linguistics is precisely to disentangle the innate parts of language from the rest (e.g., Chomsky 2005). Of course, it should also be noted that the PLN itself is not a theory of language acquisition. However, if the PLN is remotely plausible then it suggests that the UG/disentangling project is not something that could be properly expressed at the level of a linguistic theory. That is, the components of linguistic theory are themselves an irreducibly complex mixture of genetic and environmental factors.

For example, if the OCP or SSP are consequences of latching dynamics (as the PLN suggests), then they neither need to be independently learned nor innately specified, since they appear to be largely coextensive with latching dynamics. They could perhaps be equated with Chomsky's *third factor* (Chomsky 2005), however even this categorisation may be too coarse. Because although the OCP and SSP do seem to follow from a purportedly more general mechanism (i.e. latching), it is also true that these behaviours appear to depend on the way the memories themselves are encoded, which seems to be a fact about phonological inventories and the features which define them.

The SSP, for example, is dependent on the particular properties of manner features—namely that they loosely cluster the inventory into two groups along a single dimension: sonorants and obstruents. Given this clustering, latching dynamics seems to naturally produce oscillation between the two clusters. Thus, the SSP is the result of a complex interaction between something specific to phonology (sonority) and something much more general (latching dynamics). Of course, this interaction is not necessarily captured at the level of linguistic formalisms, meaning that the relevant subdivision into innate/learned/third-factor cannot occur at the level of the linguistic theory itself.

This does not necessarily entail that UG is a doomed project, merely that the complex influence of genetic and environmental factors on language acquisition may only be understandable when we integrate insights from linguistic theory into neutrally inspired models such as the PLN (and beyond, into neurobiology, etc.). Thus, properly defining UG may not be a problem that linguists can solve in isolation. This conclusion could render moot long standing discussions about the innateness of (e.g.,) phonological features (e.g., Mielke 2008), since features might not be atomic objects which can be neatly described as either innate or learned.

Of course, this brief discussion of learning is by no means exhaustive. It is intended merely to demonstrate how intermediate, neurally-inspired models such as the PLN can help to bridge the gap between linguistics and neuroscience in a way that permits more nuanced argumentation, rather than causing “interdisciplinary cross-sterilization” (Poeppel & Embick 2005). The ultimate goal is integration of linguistic and neuroscientific theories into a grander understanding of the mind/brain and, while this goal is certainly a long way off, models such as the PLN do present us with a potential way forward.

Data Availability

The data and code used for analysis in this article are available by contacting the author.

References

- Alderete, John, & Tupper, Paul. 2018. Connectionist approaches to generative phonology. In S. J. Hannahs & Anna R. K. Bosch (eds.), *The Routledge Handbook of Phonological Theory*. Routledge, New York, 360–390.
- Assaneo, Maria Florencia, & Poeppel, David. 2018. The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances* 4(2): eaao3842.
- Block, Ned. 1995. The mind as the software of the brain. In E. E. Smith & D. N. Osherson (eds.), *An Invitation to Cognitive Science*, vol. 3: *Thinking* (2nd edn.). Cambridge, MA: MIT Press.
- Boboeva, Vizhe, Brasselet, Romain, & Treves, Alessandro. 2018. The capacity for correlated semantic memories in the cortex. *Entropy* 20(11), 824.
- Bouchard, Kristopher E., Mesgarani, Nima, Johnson, Keith, & Chang, Edward F. 2013. Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495, 327–332.
- Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36(1), 1–22.
- Clements, George N. 1990. The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge, United Kingdom: Cambridge University Press, 283–333.
- Hickok, Greg, & Poeppel, David. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8, 393–402.
- Hopfield, John J. 1982. Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Sciences of the United States of America* 79, 2554–2558.
- Hopfield, John J. 1984. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America* 81, 3088–3092.

- Kang, Chol J., Naim, Michelangelo, Boboeva, Vizhe, & Treves, Alessandro. 2017. Life on the edge: Latching dynamics in a Potts neural network. *Entropy* 19, 468.
- Kanter, Ido. 1988. Potts-glass models of neural networks. *Physical Review A* 37(7), 2739–2742.
- Kaplan, Abby. 2010. *Phonology shaped by phonetics: The case of intervocalic lenition*. Santa Cruz, CA: University of California dissertation. doi:10.7282/T30G3J2K
- Krämer, Martin. 2019. Is vowel nasalisation phonological in English? A systematic review. *English Language & Linguistics* 23(2), 405–437.
- Lakoff, George. 1988. A suggestion for a linguistics with connectionist foundations. In D. Touretzky (ed.), *Proceedings of the 1988 Connectionist Summer School*. Burlington, MA: Morgan Kaufmann.
- Lashley, Karl S. 1951. *The Problem of Serial Order in Behavior*. Oxford, United Kingdom: Bobbs-Merrill.
- Marr, David & Poggio, Tomaso. 1976. From understanding computation to understanding neural circuitry. *Artificial Intelligence Laboratory. A.I. Memo*. MIT. Retrieved from <https://dspace.mit.edu/handle/1721.1/5782>
- McCarthy, John J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17, 207–263.
- Mesgarani, Nima, Cheung, Connie, Johnson, Keith, & Chang, Edward F. 2014. Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010.
- Mielke, Jeff. 2008. *The Emergence of Distinctive Features*. Oxford, United Kingdom: Oxford University Press.
- Naim, Michelangelo, Boboeva, Vizhe, Kang, Chol J., & Treves, Alessandro. 2018. Reducing a cortical network to a Potts model yields storage capacity estimates. *Journal of Statistical Mechanics: Theory and Experiment* 4, 043304.
- Nasrabadi, Nasser M., Choo, C. Y. 1992. Hopfield network for stereo vision correspondence. *IEEE Transactions on Neural Networks* 3, 5–13.
- Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. *Papers in Laboratory Phonology* 1, 258–275.
- Odden, David. 2013. Formal phonology. *Nordlyd* 40(1), 249–273.
- Pirmoradian, Sahar, & Treves, Alessandro. 2012. A talkative Potts attractor neural network welcomes BLISS words. *BMC Neuroscience* 13(Suppl. 1): P21.
- Pirmoradian, Sahar, & Treves, Alessandro. 2014. Encoding words into a Potts attractor network. In Julien Mayor & Pablo Gomez (eds.), *Computational Models of Cognitive Processes: Proceedings of the 13th Neural Computation and Psychology Workshop*. Singapore: World Scientific, 29–42.
- Poeppel, David & Embick, David. 2005. Defining the relationship between linguistics and neuroscience. In A. Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*. Mahwah, NJ: Lawrence Erlbaum, 103–118.
- Prince, Alan & Smolensky, Paul. 1997. Optimality: From neural networks to universal grammar. *Science* 275(5306), 1604–1610.
- Russo, Eleonora, & Treves, Alessandro. 2012. Cortical free-association dynamics: Distinct phases of a latching network. *Phys. Rev. E*. 85(5), 051920.

- Song, Sanming, Yao, Hongxun, & Treves, Alessandro. 2014. A modular latching chain. *Cogn Neurodyn* 8, 37–46.
- Treves, Alessandro. 2005. Frontal latching networks: A possible basis for infinite recursion. *Cognitive Neuropsychology* 22(3–4), 276–291.
- Truesdell, Clifford, & Muncaster, Robert G. 1980. *Fundamentals of Maxwell's Kinetic Theory of a Simple Monatomic Gas*. New York: Harcourt Brace Jovanovich.
- Tsodyks, Misha V., & Feigelman, Mikhail V. 1988. The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters* 6(2), 101–105.
- Walter, Mary A. 2007. Repetition avoidance in human language. Cambridge, MA: MIT doctoral dissertation.

Appendix

This appendix contains the parameters and phonological inventory used in this study.

The results in Section 4 were all obtained from simulations using a constant set of network parameters:

$$\begin{aligned}
 S &= 5 \\
 N &= 200 \\
 a_{\text{feat}} &= 0.2 \\
 a &= 0.25 \\
 p &= 1.1 \\
 q &= 0.1 \\
 \tau_1 &= 1.5 \\
 \tau_2 &= 70 \\
 \tau_3 &= 100 \\
 \beta &= 4 \\
 w &= 1.8 \\
 U &= 0.45 \\
 c_{\text{int}} &= 0.2 \\
 c_{\text{ext}} &= 0.2
 \end{aligned}$$

The inventory of phones and their featural specification is given in the table on the next page (*Table Appendix 1*). Note that the ordering of the features in the table reflects the weighting of the features within each sub-network.

Manner ↓	Place →																									
	consonantal	vocalic	sonorant	nasal	continuant	approx.	del.rel.	lateral	sibilant	dental	retroflex	apical	post-alv	alveolar	anterior	glottal	velar	coronal	dorsal	labial	exterior	front	low	high		
p	+																									
f	+				+									+												
m	+		+																		+					
t	+											+						+								
s	+				+									+				+								
ʃ	+				+							+						+				+				
tʃ	+											+						+				+				
θ	+				+								+					+								
l	+				+									+				+								
n	+		+		+								+					+								
k	+																+		+							
ŋ	+		+		+												+		+							
ʔ	+														+								+			
h	+				+											+							+			
w					+														+		+					
r					+						+							+					+			
i														+					+				+			
ɪ																							+			
e					+																		+			
a					+																		+			
æ					+																		+			
o																					+					
u																					+					


Appendix Table 1: Inventory of phones and their featural specification.

Humans Discriminate Individual Zebra Finches by Their Song

Sabrina Schalz^{1,*} & Thomas E. Dickins¹

¹ Department of Psychology, Middlesex University, London, United Kingdom

* Corresponding author: ss3903@live.mdx.ac.uk

 SS: <https://orcid.org/0000-0003-0237-4367>
TED: <https://orcid.org/0000-0002-5788-0948>

Comparative experiments have greatly advanced the field of biolinguistics in the 21st century, but so far very little research has focused on human perception of non-human animal vocalizations. Studies with zebra finch (*Taeniopygia guttata*) songs found that humans cannot perceive the full range of acoustic cues that zebra finches hear in their songs, although it remained unclear how much individual information is lost. Individual heterospecific discrimination by humans has only been shown with rhesus monkey (*Macaca mulatta*) voices. The present study examined whether human adults could discriminate two individual zebra finches by their songs, using a forced-choice Same-Different Paradigm. Results showed that adults can discriminate two individual zebra finches with high accuracy and without prior training. Discrimination mostly relied on differences in pitch contour, but discrimination was still possible with lower accuracy when pitch contour was removed. Future studies should expand these findings with more diverse non-human animal vocalizations.

Keywords: individual discrimination; zebra finch; pitch contour; human perception


1. Introduction

In the 21st century, the study of biolinguistics has made significant advances through comparative experiments with animal models. The majority of comparative studies so far have focused on non-human animals' perception of human language to draw inferences about which aspects of language are unique to humans (faculty of language in a narrow sense, or FLN) and which are not (faculty of language in a broad sense, or FLB; Hauser et al. 2002). For instance, java

Editors: Lluís Barceló-Coblijn, Universitat de les Illes Balears, Spain
Evelina Leivada, Universitat Rovira i Virgili, Spain

Received: 17 March 2020
Accepted: 11 May 2020
Published: 25 March 2021

ISSN 1450–3417


CC BY 4.0 License
© 2020 The authors

sparrows (*Padda oryzivora*) have been shown to discriminate patterns of emotional prosody in Japanese (Naoi et al. 2012). Java sparrows can also discriminate spoken English and Chinese (Watanabe et al. 2006), while the discrimination of Dutch and Japanese has been shown in cotton-top tamarin monkeys (*Saguinus Oedipus*; Ramus et al. 2000), rats (*Rattus norvegicus*; Toro et al. 2003), and large-billed crows (*Corvus macrorhynchos*; Schalz & Izawa 2020). Zebra finches (*Taeniopygia guttata*) can discriminate between familiar and unfamiliar song and speech in both English and Russian (Phillmore et al. 2017), perceive prosodic patterns in speech (Spierings & Cate 2014), as well as formant patterns in human speech and distinguish mono-syllabic words despite speaker variation (Ohms et al. 2010).

Very few studies have tested human subjects' perception of non-human animal vocalizations. Presumably both directions could be possible with features that are considered part of the FLB. We may for instance argue that zebra finches perceive prosodic patterns in speech because prosody is not unique to language (FLB), but instead also found in birdsong. That gives us two equally intriguing possibilities: Either zebra finch prosody and speech prosody are fundamentally the same (although maybe superficially different) and can be perceived bi-directionally in their entirety by both species, or they overlap at best partially, and hetero-specific perception is only feasible for one species but not the other. Vocal communication in non-human animals, although different from human language, are complex in their own right and human perception of acoustic details in these heterospecific vocalizations is far from trivial.

Studies with human infants have found that age plays a crucial role in the perception of heterospecific vocalizations. Both lemur (*Eulemur macaco flavifrons*) vocalizations and human speech but not backward speech support object categorization in three and four-months-old infants, but only speech promotes object categorization in six-months olds (Ferry et al. 2013). Exposure to primate vocalizations can extend this effect, while exposure to backward speech does not (Perszyk & Waxman 2016). Neonates initially show an equal preference for human speech and rhesus monkey (*Macaca mulatta*) vocalizations over synthetic sounds, while three-months-old infants prefer human speech over both rhesus monkey vocalizations and synthetic sounds (Vouloumanos et al. 2010). These results illustrate the effect of experience and age-related differences in human perception of heterospecific vocalizations and suggest that initial sensitivity to some hetero-specific vocalizations is lost early in life due to lack of exposure and relevance. Despite this age-related decline in perception, studies with adults are nevertheless relevant and informative. Experiments with adults have shown that zebra finches are far more sensitive to temporal fine structure than humans. When presented with forwards and backwards repetitions of single periods taken from zebra finch contact calls, which differed only in the order of temporal fine structure cues, zebra finches, unlike humans, were still able to discriminate them (Dooling & Lohr 2006), which suggests that their songs may contain acoustic details that they can perceive but that we cannot (Dooling & Prior 2017). Further experiments on human perception of heterospecific vocalizations across multiple age groups are necessary to gain a more detailed understanding of the extent of the FLB. The goal of this study is therefore to further examine whether human adults perceive individual differences in zebra finch songs in a Same-Different task. Due to their intra-

individually stereotyped songs, zebra finches are a suitable model organism for this endeavour.

Male zebra finches produce signature songs learned from a tutor early in life. These songs have two primary functions: courtship (Sossinka & Böhner 1980) and within-pair communication, for example to maintain the pair bond (D'Amelio et al. 2017) or to coordinate parental care (Boucaud et al. 2017). Songs follow an individualized, stereotyped pattern (the signature) and consist of an introduction followed by multiple motifs, which in turn consist of smaller elements. These motifs convey information about the identity of the individual, while introductory elements are fairly similar between males (Sossinka & Böhner 1980; Zann 1996). They contain both amplitude and temporal envelope cues that span up to multiple seconds, and fine structure cues of individual syllables, including amplitude, spectral and temporal cues (Dooling & Prior 2017). Recent studies have shown that zebra finches are primarily sensitive to the acoustic features contained within syllables as opposed to sequences of syllables (Lawson et al. 2018). These fine structure cues convey important information about the individual's identity, its sex and the specific call type (Prior et al. 2018).

Acoustic cues conveying information about the individual's identity are important components of the vocalizations of social animals, such as the zebra finch. Consequently, they can recognize conspecifics based on their song and any of their other calls using call-type specific signatures (Elie & Theunissen 2018). Humans, on the other hand, rely on passive voice cues and primarily discriminate each other based on fundamental frequency (perceived as pitch), followed by the frequency of the first formant (F1) for female voices and formant dispersion for male voices (Baumann & Belin 2010). There is a considerable sex difference in formant perception, as men are significantly better than women at using formant dispersion to assess the acoustic size of individual animals (Charlton et al. 2013), although it is unclear how far this difference extends into voice discrimination. Fundamental frequency also plays a major role in the voice systems of other animals, such as large-billed crows (Kondo et al. 2010). Individual discrimination and recognition is possible across species as well. Carrion crows (*Corvus corone*) have been found to discriminate familiar and unfamiliar human voices and jackdaw calls (Wascher et al. 2012). Captive cheetah (*Acinonyx jubatus*) can also discriminate between familiar and unfamiliar human voices (Leroux et al. 2018), while domestic dogs and domestic cats discriminate their owner's voice from that of an unfamiliar person (Adachi et al. 2007; Saito & Shinozuka 2013) and rhesus monkeys match a familiar human voice to the corresponding face (Sliwa et al. 2011). In turn, human infants (and to some degree, adults) can discriminate two individual rhesus monkeys by their voices (Friendly et al. 2014). At an age of six months, infants showed a more accurate discrimination compared to infants tested at 12 months, although with practice the 12 months old infants were able to outperform the six-month olds (Friendly et al. 2013).

The present experiment extends these findings by testing human adults' discrimination of two zebra finches by their song. As discussed above, human perception of zebra finch songs is likely far less detailed than that of zebra finches, at least with regards to temporal fine structure. The primary aim of this study is to examine whether humans are at all able to perceive individual differences in the

songs of zebra finches, and if so, to what degree. Further attention is given to explore which acoustic cues in zebra finch songs humans can use for this task, whether there is a correlation between the listener's sex and discrimination accuracy, and whether the discrimination improves with practice.

Results will extend findings on humans' perception of zebra finch songs, and more generally offer further insights into the commonalities between human and non-human vocalizations.

2. Material and Methods

The study was separated into condition 1 with natural zebra finch songs and condition 2 with manipulated songs as described in section 2.2. Condition 2 was designed as an extension to the previously conducted condition 1, which is reflected in its smaller sample size and analysis. Apparatus and procedure were the same for both conditions. The analysis was mostly the same unless stated otherwise for the respective aspect. Results were analysed in three parts to address the core questions: whether humans can discriminate individual zebra finches by their song and if so, how accurately, which acoustic cues play a role in this discrimination, and whether discrimination accuracy improves over time. Both conditions were approved by the Middlesex University Psychology Research Ethics Committee.

2.1. Participants

Participants were 50 adults (25 female) in condition 1, and 25 adults (14 female) in condition 2. All were students and staff at Middlesex University between the ages 18 to 50. Participants did not report hearing problems and gave informed consent. No participants were removed from the analysis.

2.2. Stimuli

Stimuli in condition 1 consisted of the natural song of two male zebra finches (3 and 4 months old) recorded at Bielefeld University. Animal housing and song recording were in compliance with all applicable national guidelines for the care and use of animals. The recordings were analysed in Praat version 6.0.49 (industry standard software for acoustic analysis; Boersma & Weenink 2019) and nine motifs per individual were selected. Selection was based on high similarity in pitch contour, intensity contour, duration and number of repeated elements. Each selected motif was then high-pass filtered at 500 Hz with Audacity version 2.3.0 (<https://www.audacityteam.org>) to reduce low-frequency background noise (e.g. perch clanging against the cage bar) without influencing the high-frequency song. Motifs of zebra finch *B* were shorter than those produced by zebra finch *A*, and so recordings from *A* had to be cut to remove total duration as a possible discrimination cue. Cuts were made at element boundaries for clean breaks, and as such stimuli differed in mean duration by 0.04 s, which we considered acceptable (see *Table 1* for mean values of acoustic features). In addition to differences in pitch and formant frequencies, the motifs also differed structurally as motifs *A*

Acoustic feature	Zebra finch A mean	Zebra finch A SD	Zebra finch B mean	Zebra finch B SD
Duration per motif (ms)	397.6	10	335.5	8
Intensity per motif (dB)	59.7	0.5	59.1	1.5
Pitch per motif (Hz)	3177.4	204	2888.8	309.3
Frequency of F1 (Hz)	3183.4	33.2	3368.5	73.5
Frequency of F2 (Hz)	4743.6	53.6	5177.4	91.9
F1-F2 dispersion (Hz)	1560.6	38.4	1807	54.7

Table 1: Acoustic features of the nine motifs of each zebra finch. Frequency range was set to a minimum 50 Hz and maximum 10,000 Hz for the pitch analysis (note that Praat measures pitch instead of F0) and to a maximum 10,000 Hz and 3 extracted formants for the formant analysis.

consisted of two elements while motifs *B* consisted of three. A silent 2 s interval was added at the end of each motif to create clear breaks between them. As indicated by the spectrogram of zebra finch *A*, three formants were initially extracted but since only two were reliably found for zebra finch *B*, F3 was not further analysed in this study (see Appendix, *Figures A1* and *A2*).

Stimuli for condition 2 were taken from condition 1 and then manipulated in Praat (Boersma & Weenink 2019). To test the influence of the signature encoded in the envelope of the song on the discrimination accuracy participants achieve, pitch contour (the pitch pattern across the entire motif) was removed from the recordings. All existing pitch points were removed, and new pitch points were added at the time points 0.0001 s, 0.1 s, 0.2 s, 0.3 s, and 0.4 s at the frequency of the mean pitch of the respective stimulus. This was done to continue to include mean pitch as possible discrimination cue (see *Table 2* for resulting acoustic features). After initial manipulation, each recording was then checked, and additional pitch points were added where necessary (see Appendix *Figures A3* and *A4* for natural and manipulated pitch contour).

2.3. Apparatus

The participant background questionnaire and the discrimination task were presented in the software PsychoPy version 3.2 (Peirce et al. 2019) on a desktop computer. The experiment was conducted in a quiet room, and stimuli were played using over-ear headphones.

Acoustic feature	Zebra finch A mean	Zebra finch A SD	Zebra finch B mean	Zebra finch B SD
Duration per motif (ms)	397.6	10	335.5	8
Intensity per motif (dB)	59.7	0.5	59.1	1.5
Pitch per motif (Hz)	3124	192.6	2872.5	304.7
Frequency of the first formant (Hz)	3190.2	171.4	2962.7	222
Frequency of the second formant (Hz)	5517.7	223.5	5563.1	286.6

Table 2: Acoustic features of the nine manipulated motifs of each zebra finch. Frequency range was set to a minimum 50 Hz and a maximum 10,000 Hz for the pitch analysis and to a maximum 10,000 Hz and 3 extracted formants (indicated by the spectrograms) for the formant analysis.

2.4. Procedure

Participants were tested with the forced-choice Same-Different Paradigm (Pisoni & Lazarus 1974). Each of 40 trials contained two vocalizations combined at random to avoid predictability, either produced by the same individual (“same”-trial) or two different individuals (“different”-trial). Before each experiment, participants received verbal instructions about the discrimination task emphasising that the choice would be between individuals of the same species, not two different species. After the verbal explanation, participants were shown the following instructions on the screen reiterating the verbal instructions: “You will now hear 40 sound pairs. A pair of sounds was either produced by the same animal or by two animals of the same species. After each pair, you will be asked to decide whether you heard the same animal or two different animals. Sounds are separated by a 2 s interval and only 0.3 s long.”. Following the playback of each pair, participants were asked whether the song was sung by the same bird (keypress “y” for yes) or not (“n” for no). During the experiment, participants did not receive feedback on their discrimination accuracy.

2.5. Analysis

The analysis was conducted entirely in R (R Core Team 2019). The first part of the analysis focused on the degree of discrimination accuracy. Following the signal detection theory (Stanislaw & Todorov 1999), responses were divided into the four categories hit (y on a “same”-trial), miss (n on a “same”-trial), correct reject (n on a “different”-trial), and false alarm (y on a “different”-trial) to determine the hit rate (proportion of hit responses in same-trials) and the false alarm rate (false alarm responses in different-trials). These two values ranging from 0 to 1 were used to calculate the discrimination sensitivity index d' using the R package `psyphy` and the formula `dprime.SD(H, FA, method = "diff")`, (Knoblauch 2014). We chose d' scores over other success measures, such as the percentage of correct trials,

because they are less susceptible to participants' response biases (Stanislaw & Todorov 1999). If a participant answers "yes" in every trial (an extreme response bias), the hit rate and the false alarm rate will both be 1 and the d' score for equal rates is 0. This score reflects that the participant did not discriminate between "same"-trials and "different"-trials, whereas the percentage of correct trials depends entirely on how many "same"-trials were randomly chosen, resulting in a discrimination accuracy that could be anywhere between 0% and 100%. Since d' scores cannot be calculated with absolute values, the formula described by (Snodgrass & Corwin 1988) was used to correct absolute rates of 0 and 1 (see formula 1). Seven rates of 1 and 12 rates of 0 were corrected in condition 1, as well as 1 rate of 0 in condition 2. The lowest possible d' score of 0 was given for equal hit and false alarm rates (e.g. when a participant answers yes on every trial), and when the false alarm rate was higher than the hit rate, as d' scores cannot be negative. Consequently, d' scores ranged from 0 (no discrimination) to 5.94 (perfect discrimination) and indicate discrimination accuracy on a continuous scale rather than binary success or failure. Three single trials in condition 1 were missing and thus not included in the analysis.

$$(1) \quad \text{corrected rate} = \frac{0.5 + \text{response rate (either hit or false alarm)}}{1 + \text{number of trials (either same or different)}}$$

Non-parametric statistical tests were chosen for data which were not normally distributed based on a Shapiro-Wilk normality test. This was the case with d' scores in both conditions, the responses types in condition 1, and the success trend in condition 2. Homogeneity of variance was confirmed with a Levene test for the one-way ANOVA, using the R package "car" (Fox & Weisberg 2019).

A one-sample Wilcoxon signed-rank test was used to assess whether d' scores were significantly above chance level ($\mu = 0$). Additionally, a Mann-Whitney-U test was used to determine whether d' scores differed significantly between male and female participants. No participants were excluded from this part of the analysis.

The second part of the analysis focused on the relevance of different acoustic cues: mean pitch, mean F1, and formant dispersion (F1-F2) in Hz. These cues were chosen because they are the most important cues in human voice discrimination (Baumann & Belin 2010). As formant dispersion was very irregular in condition 2, this cue was only analysed for condition 1. If a given cue was relevant for the discrimination, "same" pairs with high differences should trigger the mistake "miss" more often, and "different" pairs with low differences should trigger the mistake "false alarm" more often. Stimuli pairs with a minimum occurrence per response type were chosen to focus on the most difficult combinations and to exclude those that only triggered the same response once or twice. For condition 1, pairs that triggered a "false alarm" or a "miss" response at least three times were selected. For condition 2, pairs that triggered a "false alarm" or a "miss" response at least four times, as well as pairs that triggered a "hit" or correct reject" at least five times were selected. These different thresholds were chosen in order to only include the most frequently occurring pairs while still including enough pairs for

analysis. “Hit” and “correct reject” responses from condition 1 were not included as the success rate was so high that the analysis of correctly categorized pairs would not be very insightful. The success rate in condition 2 was lower and the sample size smaller, which is why all four response types are included. A total of 23 “false alarm” and 22 “miss” pairs were selected for condition 1. For condition 2, 14 “false alarm”, 16 “miss”, 18 “hit”, and 17 “miss” pairs were selected. Pairs with opposite stimuli order (e.g. a2b3 and b3a2) were treated as the same pair. Every selected pair was weighted once in the acoustic cue analysis. In condition 1, acoustic parameters were compared between “false alarm” and “miss” pairs using a Mann-Whitney U test. In condition 2, a one-way ANOVA was used to analyse all four response types. Participants with a d' score of 0 were excluded from this part of the analysis since they did not perceive any difference between stimuli (one excluded in condition 1, four in condition 2).

The third part of the analysis focused on the discrimination accuracy over time. A trend in discrimination success (measured as percentage of correct answers pooled from all participants per condition) was analysed with a linear regression model $\text{lm}(\text{percentage correct} \sim \text{trial number})$ for condition 1, and a Mann-Kendall trend test for condition 2 using the R package “Kendall” (McLeod 2011). No participants were excluded from this part of the analysis.

3. Results

In condition 1, the average d' score was 3.68 (SD = 1.54, 95% CI [3.24, 4.11]) with individual scores ranging from 0 to 5.94, the highest possible score. In condition 2, the average d' score was 1.3 (SD = 0.82, 95% CI [0.96, 1.63]) and individual scores ranged from 0 to 3.29. d' scores in both conditions were significantly above chance level ($p < 0.01$), and d' scores in condition 2 were significantly below scores from condition 1 ($p < 0.01$; see *Figure 1*). There was no significant difference in d' scores between female and male participants in either condition.

Neither mean pitch nor mean F1 or mean formant dispersion were significantly lower in “false alarm” responses than “miss” responses in condition 1.

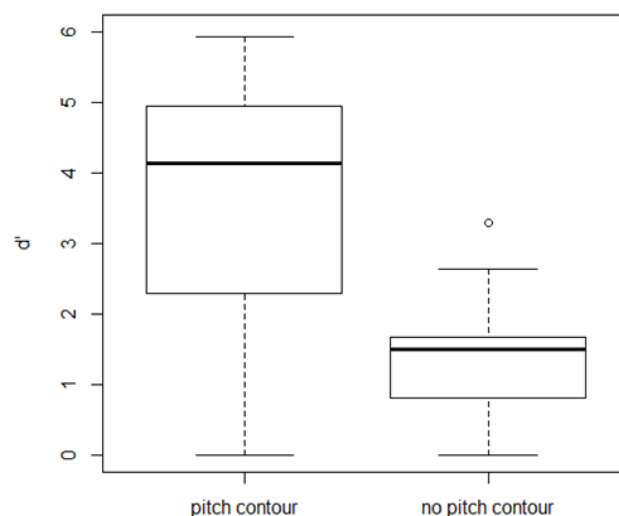


Figure 1: D' scores obtained in condition 1 (with pitch contour) and condition 2 (without pitch contour).

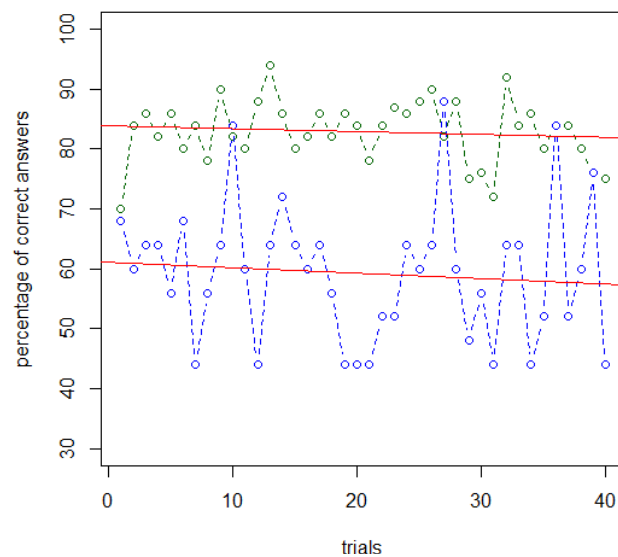


Figure 2: Percentage of correct answers (either “hit” or “correct reject”) for each trial pooled from all participants per condition. Condition 1 is drawn in green, condition 2 in blue. The red linear regression lines indicate the overall trend for each condition.

There were also no significant differences in mean pitch or mean F1 frequency between response types in condition 2.

There was no significant trend across trials in either condition ($m = -0.04$ in condition 1 and $\tau = -0.12$ in condition 2; see Figure 2).

4. Discussion

These results show that humans can discriminate two individual zebra finches based on a short section of their song, even if pitch contour is not available as a discrimination cue. Overall, discrimination accuracy was very high with the majority of participants reaching either perfect or high d' scores, although success was highly variable inter-individually (see Figure 1).

This level of discrimination accuracy is especially remarkable since humans likely cannot perceive all details in zebra finch songs (Dooling & Prior 2017). While discrimination was far from perfect and there are surely some acoustic cues that participants did not perceive, this study shows that those cues that we do perceive are still enough for reliable individual discrimination. The most salient cue for this discrimination task seems to be pitch contour, a temporal envelope cue. Scores obtained in condition 2 without pitch contour were significantly lower and the percentage of correct answers across trials was more than 20 % lower in condition 2 compared to condition 1. The primary cue being part of the song envelope is in accordance with previous findings that humans are relatively insensitive to fine structure cues (Dooling & Lohr 2006).

This also suggests that our cue weighting of zebra finch songs differs considerably from that of zebra finches who are relatively insensitive to syllable sequences and instead focus on fine structure within syllables (Lawson et al. 2018). Since discrimination was still possible in condition 2 despite the removal of this cue, there must also be other, albeit less important cues that participants perceived

additionally. The analysis of stimuli pairs that triggered certain response types found that mean pitch, mean F1, and mean formant dispersion frequencies are unlikely to be contributing cues. This is contrary to findings that these three features are the most important cues for humans in voice discrimination (Baumann & Belin 2010). Mean pitch frequencies showed some variation intra-individually, but F1 and formant dispersion frequencies were fairly stereotyped between the two individuals (see *Tables 1* and *2*) and would have been available as useful cues.

Additionally, there was no difference in discrimination success between men and women, which has been observed for formant perception in acoustic size judgements (Charlton et al. 2013). Consequently, formants do not seem to be relevant for this task, although it is currently unclear why. Mean amplitude and overall duration were not available as cues, since they were standardized for all stimuli. By exclusion this leaves amplitude contour, timbre, and possibly, to some superficial degree, fine structure as possible cues in condition 2, and their potential relevance should be explored in future experiments. However, it is possible that the acoustic cues used by participants also vary inter-individually. Relevance of acoustic cues was analysed at the group level, but selective attention to certain cues over others and employed perceptual strategies could differ between individuals (Holt et al. 2018). Additionally, differences in participants' backgrounds (such as tonal languages or music training) may contribute to further attentional biases. Much more work is needed to narrow in on the acoustic cues that humans extract from zebra finch songs and how these may vary between different individuals and backgrounds.

The trend analysis (see *Figure 2*) shows that discrimination success is already high in the first trials without prior training. This is contrary to expectations based on previous findings on infants' sensitivity to non-human primate vocalizations that showed a rapid decrease in sensitivity with age and lack of exposure (Ferry et al. 2013; Perszyk & Waxman 2016; Vouloumanos et al. 2010). Even more so, it is contrary to the findings from the discrimination experiment with rhesus monkey voices in which adults only achieved an average d' score of 0.37 (Friendly et al. 2014), which is far below the mean d' scores of 3.68 and 1.3 observed here. To a large extent, this is likely due to the signature component of zebra finch songs, which is possibly easier to perceive than passive voice cues. Still, adults in condition 2 still outperformed those in the rhesus monkey study and it would be worth exploring how the discrimination of other animals would compare to these scores. The trend analysis also shows that participants' discrimination accuracy did not improve with practice, although accuracy could potentially increase with more extensive exposure exceeding 40 trials. However, the a priori high discrimination accuracy and lack of significant improvement show that this task does not require previous exposure or explicit training.

5. Conclusion

This study has shown that human adults are very sensitive to individual differences in zebra finch songs and predominantly use pitch contour to discriminate two individuals, although other acoustic cues play a role as well. Human participants do not seem to rely on mean pitch or mean formant frequencies in this

discrimination task. Discrimination accuracy is high without prior training and far exceeds the discrimination abilities observed for rhesus monkey voices in adults.

In the 21st century, the field of biolinguistics has made great advances in our understanding of shared features in human speech through comparative studies on non-human animals' perception of language, but the results obtained here show that we have not yet reached the limitations of our own perceptual capabilities with regards to heterospecific vocalizations. Going forward, more work should focus on exploring which components of non-human animal vocalizations humans of all age groups can perceive, which acoustic cues are used for this perception, and most intriguingly, why they can be perceived across species in the first place.

Data Availability

The data and code used for the analysis in this article are freely available from Figshare: <https://doi.org/10.6084/m9.figshare.c.4998065.v1>

Author Contributions

S.S. designed and performed the experiments, conducted the analysis, and led the writing of the manuscript. T.D. supervised the project, added to the argument, and contributed to the draft.

Acknowledgements

We would like to thank Barbara Caspers for enabling the recording of the zebra finch songs.

Appendix

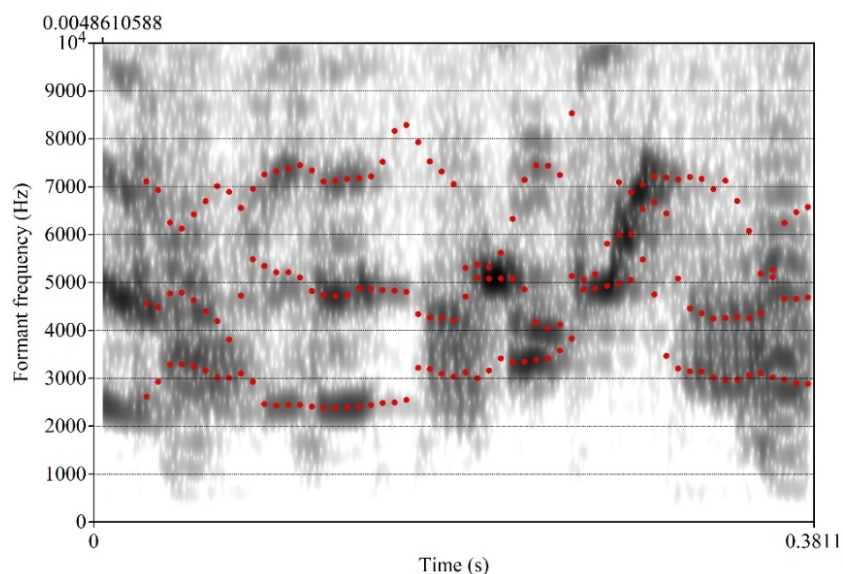


Figure A1: Sample spectrogram ranging from 0 Hz to 10,000 Hz showing one motif of zebra finch A with extracted formants drawn in (red dots) obtained in Praat (Boersma & Weenink 2019). Light area indicates pause between the two elements.

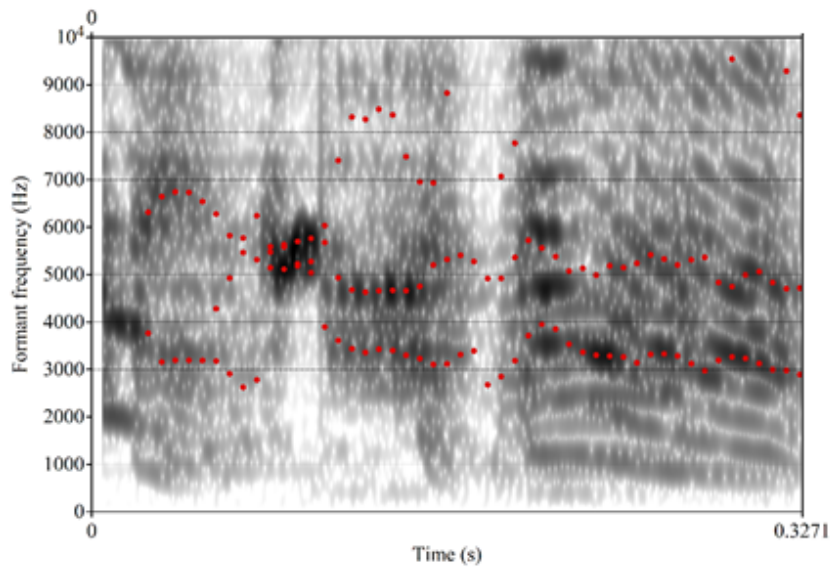


Figure A2: Sample spectrogram ranging from 0 Hz to 10,000 Hz showing one motif of zebra finch B with extracted formants drawn in (red dots) obtained in Praat (Boersma & Weenink 2019). Light areas indicate pauses between elements.

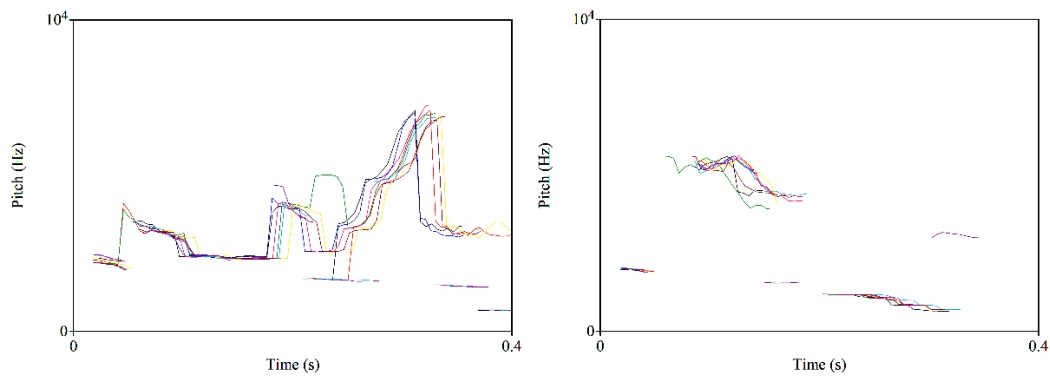


Figure A3: Natural pitch contour of the stimuli used in condition 1 produced by zebra finch A (left) and zebra finch B (right). Each colour corresponds to one motif per zebra finch.

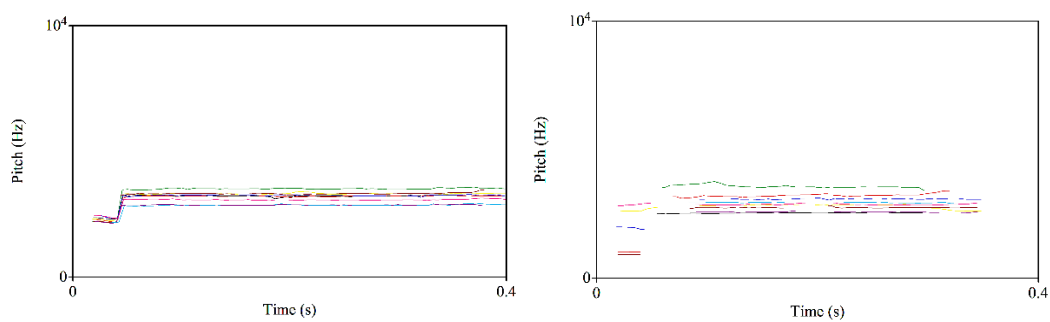


Figure A4: Manipulated pitch contour of the stimuli used in condition 2 produced by zebra finch A (left) and zebra finch B (right) where pitch contour was equalized. Each colour corresponds to one motif per zebra finch.

References

- Adachi, Ikuma, Hiroko Kuwahata & Kazuo Fujita. 2007. Dogs recall their owner's face upon hearing the owner's voice. *Animal Cognition* 10(1), 17–21.
- Baumann, Oliver & Pascal Belin. 2010. Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research* 74(1), 110–120.
- Boersma, Paul & David Weenink. 2019. *Praat: Doing Phonetics by Computer*. <http://www.fon.hum.uva.nl/praat>.
- Boucaud, Ingrid C.A., Emilie C. Perez, Lauriane S. Ramos, Simon C. Griffith & Clémentine Vignal. 2017. Acoustic communication in zebra finches signals when mates will take turns with parental duties. *Behavioral Ecology* 28(3), 645–656.
- Charlton, Benjamin D., Anna M. Taylor & David Reby. 2013. Are men better than women at acoustic size judgements? *Biology Letters* 9(4): 20130270, doi: [10.1098/rsbl.2013.0270](https://doi.org/10.1098/rsbl.2013.0270).
- D'Amelio, Pietro B., Lisa Trost & Andries Ter Maat. 2017. Vocal exchanges during pair formation and maintenance in the zebra finch (*taeniopygia guttata*). *Frontiers in Zoology* 14: 13, doi: [10.1186/s12983-017-0197-x](https://doi.org/10.1186/s12983-017-0197-x).
- Dooling, Robert J. & Bernard Lohr. 2006. Auditory temporal resolution in the zebra finch (*taeniopygia guttata*): A model of enhanced temporal acuity. *Ornithological Science* 5(1), 15–22.
- Dooling, Robert J. & Nora H. Prior. 2017. Do we hear what birds hear in birdsong? *Animal Behaviour* 124, 283–289.
- Elie, Julie E. & Frédéric E. Theunissen. 2018. Zebra finches identify individuals using vocal signatures unique to each call type. *Nature Communications* 9(1), 1–11.
- Ferry, Alissa L., Susan J. Hespos & Sandra R. Waxman. 2013. Nonhuman primate vocalizations support categorization in very young human infants. *PNAS* 110(38), 15231–15235.
- Fox, John & Sanford Weisberg. 2019. *car: Companion to Applied Regression*. R package version 3.0-5, <https://cran.r-project.org/package=car>.
- Friendly, Rayna H., Drew Rendall & Laurel J. Trainor. 2014. Learning to differentiate individuals by their voices: Infants' individuation of native- and foreign-species voices. *Developmental Psychobiology* 56(2), 228–237.
- Hauser, Marc D., Noam Chomsky & W. T. Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598), 1569–1579.
- Holt, Lori L., Adam T. Tierney, Giada Guerra, Aeron Laffere & Frederic Dick. 2018. Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research* 366, 50–64.
- Knoblauch, Kenneth. 2014. *psyphy: Functions for analyzing psychophysical data in R*. R package version 0.1-9, <https://cran.r-project.org/package=psyphy>.
- Kondo, Noriko, Ei-Ichi Izawa & Shigeru Watanabe. 2010. Perceptual mechanism for vocal individual recognition in jungle crows (*Corvus macrorhynchos*): contact call signature and discrimination. *Behaviour* 147(8), 1051–1072.
- Lawson, Shelby L., Adam R. Fishbein, Nora H. Prior, Gregory F. Ball & Robert J. Dooling. 2018. Relative salience of syllable structure and syllable order in

- zebra finch song. *Animal Cognition* 21(4), 467–480.
- Leroux, Maël, Robyn S. Hetem, Martine Hausberger & Alban Lemasson. 2018. Cheetahs discriminate familiar and unfamiliar human voices. *Scientific Reports* 8(1), 1–6.
- McLeod, A. I. 2011. *Kendall: Kendall rank correlation and Mann-Kendall trend test*. R package version, 2.2 <https://cran.r-project.org/package=Kendall>.
- Naoi, Nozomi, Shigeru Watanabe, Kikuo Maekawa & Junko Hibiya. 2012. Prosody discrimination by songbirds (Padda oryzivora). *PLoS ONE* 7(10): e47446.
- Ohms, Verena R., Arike Gill, Caroline A. A. van Heijningen, Gabriel J. L. Beckers & Carel ten Cate. 2010. Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society B* 277(1684), 1003–1009.
- Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas K. Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1), 195–203.
- Perszyk, Danielle R. & Sandra R. Waxman. 2016. Listening to the calls of the wild: The role of experience in linking language and cognition in young infants. *Cognition* 153, 175–181.
- Phillmore, Leslie S., Jordan Fisk, Simone Falk & Christine D. Tsang. 2017. Songbirds as objective listeners: Zebra finches (*Taeniopygia guttata*) can discriminate infant-directed song and speech in two languages. *International Journal of Comparative Psychology* 30: 32722.
- Pisoni, David B. & Joan H. Lazarus. 1974. Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America* 55(2), 328–333.
- Prior, Nora H., Edward Smith, Shelby Lawson, Gregory F. Ball & Robert J. Dooling. 2018. Acoustic fine structure may encode biologically relevant information for zebra finches. *Scientific Reports* 8(1): 6212.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ramus, F., M. D. Hauser, C. Miller, D. Morris & J. Mehler. 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288(5464), 349–351.
- Saito, Atsuko & Kazutaka Shinozuka. 2013. Vocal recognition of owners by domestic cats (*felis catus*). *Animal Cognition* 16(4), 685–690.
- Schalz, Sabrina & Ei-Ichi Izawa. 2020. Language discrimination by large-billed crows. *The Evolution of Language (EvoLang13): Proceedings of the 13th International Conference*, 380–389. https://brussels.evolang.org/proceedings/evolang13_proceedings.pdf.
- Sliwa, Julia, Jean-René Duhamel, Olivier Pascalis & Sylvia Wirth. 2011. Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *PNAS* 108(4), 1735–1740.
- Snodgrass, Joan G. & June Corwin. 1988. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General* 117(1), 34–50.
- Sossinka, Roland & Jörg Böhner. 1980. Song types in the zebra finch poephila

- guttata castanotis1. *Zeitschrift für Tierpsychologie* 53(2), 123–132.
- Spierings, Michelle J. & Carel ten Cate. 2014. Zebra finches are sensitive to prosodic features of human speech. *Proceedings of the Royal Society B* 281(1787): 20140280.
- Stanislaw, H. & N. Todorov. 1999. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers* 31(1), 137–149.
- Toro, Juan M., Josep B. Trobalon & Núria Sebastián-Gallés. 2003. The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition* 6(2), 131–136.
- Vouloumanos, Athena, Marc D. Hauser, Janet F. Werker & Alia Martin. 2010. The tuning of human neonates' preference for speech. *Child Development* 81(2), 517–527.
- Wascher, Claudia A. F., Georgine Szipl, Markus Boeckle & Anna Wilkinson. 2012. You sound familiar: Carrion crows can differentiate between the calls of known and unknown heterospecifics. *Animal Cognition* 15(5), 1015–1019.
- Watanabe, Shigeru, Erico Yamamoto & Midori Uozumi. 2006. Language discrimination by Java sparrows. *Behavioural Processes* 73(1), 114–116.
- Zann, Richard A. 1996. *The Zebra Finch: A Synthesis of Field and Laboratory Studies* (Oxford ornithology series 5). Oxford: Oxford University Press.

In Defence of FLB/FLN: A Reply to Wacewicz et al. (2020)

José-Luis Mendívil-Giró^{1,*}

¹ Department of Linguistics, University of Zaragoza, Spain

* Correspondence: jlmendi@unizar.es

 JLMG: <https://orcid.org/0000-0001-6891-6275>

1. Introduction

“*Language* in language evolution research” (Wacewicz et al., 2020) is a valuable synthesis of recent research on the origin and evolution of language. Its central message, however, is surprising: instead of encouraging the clearest possible specification of the object of study in research on language evolution, that is, a specification of what evolves when we say that language evolved, the authors argue that the lack of agreement on what language is has served to facilitate the significant progress made in research on language evolution over the last decade.

The absence of a more or less hegemonic theory of language (within and outside linguistics) may have made easier the proliferation of hypotheses, data sources, methodologies and opportunities for interdisciplinary collaboration in language evolution research. But, unless one wants to say that a certain definition of language can be an obstacle to the investigation of its evolution, it is difficult to justify the claim that things would have been worse, or progress less marked, had there been more agreement on the nature of the object of study. It is more logical to suppose that things might have gone even better, although it is impossible to know.

The argument Wacewicz et al. (2020) use to show that a precise definition of *language* in the expression *language evolution* is unnecessary is based on the fact that the best known attempt to establish a terminology intended to favour the collaboration between disciplines and schools of linguistic thought in the field of the study of language evolution has been a failure, and that such a failure has not prevented these studies from flourishing. The authors refer, of course, to the terminological proposal of Hauser et al. (2002; henceforth HCF) to distinguish the

Editor: Lluís Barceló-Coblijn, Universitat de les Illes Balears, Spain
Evelina Leivada, Universitat Rovira i Virgili, Spain

Received: 27 October 2020
Accepted: 2 November 2020
Published: 25 March 2021

ISSN 1450–3417



CC BY 4.0 License
© 2020 The author

faculty of language in the broad sense (FLB) and the faculty of language in the narrow sense (FLN). The reason why they suggest that the HCF proposal is a failure is that Chomsky, Fitch and Hauser would have defined the notion of FLN in a contradictory and inconsistent way in HCF and in Fitch et al. (2005; henceforth FHC). As argued in Wacewicz et al. (2020), what is relevant is that this incoherence has gone unnoticed in the field (except for Wacewicz et al. themselves), which would prove the irrelevance of precise definitions of the object of study in the field of language evolution research.

It would certainly be naive to think that we can have a “correct” definition of language. The same is true in any field of science: you do not need a universally accepted definition of life to study the origin of life, nor a universally accepted definition of natural species to study the origin of species (not to mention matter or energy). Yet shared assumptions about these objects are clearly needed if the sciences that study them are to be viable. In the absence of such assumptions, one cannot speak of biology or evolutionary theory in any grounded or meaningful way. There is no reason why (evolutionary) linguistics should be different here.

However, Wacewicz et al.’s central message seems to be that the (alleged) failure of the FLB/FLN terminology would have “freed” the study of language evolution from a restrictive conception of language, thus facilitating the proliferation of new points of view and new opportunities for interdisciplinary collaboration.

As I hope to show, it is not true that the definition of FLN is contradictory or inconsistent, nor is it true that terminological proposals, such as that of HCF, are irrelevant or unnecessary for the evaluation of real progress in research on language evolution.

2. On the Definition(s) of FLN

Wacewicz et al. imply that the HCF proposal intends to define language as FLN, that is, that HCF seeks to determine the FLN as the “authentic” object of study of language evolution research, rather than the whole FLB. But this is not so at all. In reality, HCF is not about language in general, but about the human faculty of language (FLB), a property or state of the human brain that allows us to learn and use languages. More specifically, HCF is a proposal on “unpacking FLB into its myriad component mechanisms” (Fitch et al. 2005: 181) with the aim of better understanding how these different components have been able to evolve, under the reasonable assumption that “profitable research into the biology and evolution of language requires fractionation of ‘language’ into component mechanisms and interfaces” (Fitch et al. 2005: 179). Whereas we can appreciate that the choice of the adjectives *broad* and *narrow* respond to a certain conception of language, in no way can it be suggested that in both articles *language* is defined as FLN, nor that it is implied that the study of the evolution of language should be limited to the study of FLN. In fact, the object of study from the biolinguistic point of view adopted by HCF cannot be other than FLB (that is, FL), and one of the aspects of this study is to determine what components it has, and whether FLN (a subset of FLB) really exists as such or not. Of course, it is true that

Chomsky and others have continued to use the word *language* to refer to this specific part of the FLB, instead of using the expression FLN, although Chomsky has also used *computational system* (Chomsky 1995) and *basic property* (Chomsky 2016). Moreover, it is debatable whether something that does not include the sensorimotor component should still be called *language* (albeit in the narrow sense). In any case, we might note that using the word *language* to designate any of the components of language is not a sin exclusive to generative grammar.

Besides, it is not true, as stated in Wacewicz et al. (2020), that HCF and FHC offer two different definitions of FLN, and that they are incompatible:

The proponents of FLN have defined it twice, in fundamentally discrepant ways: the definition originally formulated in Hauser et al. (2002) and the one later formulated in Fitch et al. (2005) describe two distinct entities (not simply different versions of the same entity).

(Wacewicz et al., 2020: 64)

In HCF, Hauser, Chomsky and Fitch propose that the FLN label should be reserved, by convention, for those components of the FL that (supposedly) are neither shared with other species (are specifically human) nor are part of other human cognitive domains (are language-specific); hence the use of the word *narrow*. Their hypothesis in HCF is that the FLN label should be reserved only for the computational component: “We propose in this hypothesis that FLN comprises only the core computational mechanisms of recursion as they appear in narrow syntax and the mapping to the interfaces” (Hauser et al. 2002: 1573), assuming, therefore, that only those elements meet the requirements to be included in that category.

What other definition of FLN appears in FHC that implies a different entity as a referent for FLN, or is in contradiction with the definition of HCF? None.

HCF reserves the term FLN for the computational component because its authors believe (as a hypothesis) that it is the only component of FLB that is specifically human and specifically linguistic. In FHC there is no change of reference or definition: the authors continue to maintain the same content of FLN. In fact, they discuss in detail why they think that the FL components that Pinker & Jackendoff (2005) (an article to which FHC is a response) also consider as specifically linguistic and human (certain aspects of human speech) are not part of FLN, but of FLB.

The supposed contradiction pointed out in Wacewicz et al. (2020) can be explained if we consider that the definition of FLN in HCF is *extensional* while the definition in FHC is *intensional*, but there is no inconsistency or contradiction. Actually, FHC alludes to this fact: “The term ‘FLN’ thus served dual duties in HCF” (Fitch et al. 2005: 182). Note that the intensional use of the term FLN is not new in FHC. In HCF the authors use it when commenting on Liberman’s approach:

For example, Liberman and his associates [...] have argued that the sensory-motor systems were specifically adapted for language, *and*

hence should be considered part of FLN.

(Hauser et al. 2002: 1569, emphasis added)

In both texts there is the same terminological proposal ('let us call FLN that which is specific to human language') and the same empirical hypothesis ('only the computational system is specific to human language, that is, FLN'). Why does HCF consider that the FLN is made up of the computational component? Because Hauser, Chomsky and Fitch stipulate that only the computational component is specific to human language, but not because there is some other inherent connection between the adjective *narrow* and the computational component of FL. If the definition criterion of FLN in HCF were not (human and linguistic) specificity, it would make no sense that other components of FL as different as the conceptual-intentional (CI) and the sensorimotor (SM) systems were grouped into what is not FLN. What groups CI and SM together against FLN is that they are (by hypothesis) neither specifically human nor specifically linguistic. Hence, they are part of FLB, but not of the FLN subset. So, the reasoning seems to be as follows: we think that the computational system is specific to human language; we want to call that which is specific to human language FLN; hence, the FLN is the computational system.

Wacewicz et al. (2020) illustrate their argument by citing the following texts as an example of contradiction:

The contents of FLN are to be empirically determined, and could possibly be empty, if empirical findings showed that none of the mechanisms involved are uniquely human or unique to language, and that only the way they are integrated is specific to human language. The distinction itself is intended as a terminological aid to interdisciplinary discussion and rapprochement, and obviously does not constitute a testable hypothesis. (Fitch et al. 2005: 181)

Second, although we have argued that most if not all of FLB is shared with other species, whereas FLN may be unique to humans, this represents a tentative, testable hypothesis in need of further empirical investigation. (Hauser et al. 2002: 1576)

But in both texts Chomsky, Fitch and Hauser indicate that the content of FLN must be empirically determined (and that it is possible that it is an empty set). The statement in the text from 2002 that the FLN (extensionally identified as the computational component) is specific to human language is a hypothesis in need of further investigation; the naming of both sets of components as broad or narrow (which is what the end of the text from 2005 refers to) is obviously not a falsifiable hypothesis, but a terminological convention. There is no contradiction.

Wacewicz et al. point out that "[i]nterestingly, Hauser, Chomsky and Fitch themselves have never addressed the inconsistency and may even remain unaware of it" (2020: 66), which is not surprising if such inconsistency only exists in their own interpretation. Wacewicz et al. (2020: 68) affirm that the important

thing is not the possible inconsistency, but the fact that it went unnoticed, which would supposedly be an argument in favour of the irrelevance of “top-down” definitions of the object of study. The argument (insofar as it has any force) is not actually evaluable, because the absence of the perception of incongruity is easily explained considering that there is no incongruity.

Moreover, the influence that the terminological proposal has had in the field could be estimated, for example, through considering the number of citations of HCF, which, according to Wacewicz et al., “is doubtlessly among the most influential works in the field, and probably its most widely cited article” (2020: 63).

3. Applying the HCF Terminology: Speech, Communication, and the Language-Ready Brain

The remainder of the article is a comprehensive and informative synthesis of the last ten years of research in language evolution, mostly that which does not use the HCF terminology. But since FLB/FLN is a terminological proposition, and not a theory of language, it is easy to see that it would have been very useful (or, at any rate, perfectly possible) to use it to classify and explain the four main lines of research (or models) that the authors consider, which are the following ones (including the main authors of each):

- (i) Language as a Multimodal Phenomenon (Kendon, McNeill, Zlatev)
- (ii) Language as a Complex Adaptive System (Steels, Kirby)
- (iii) Language as a form of social interaction (Tomasello, Levinson)
- (iv) Language in the Language-Ready Brain (Arbib, Bouchard, Boeckx and Benítez-Burraco)

Model (i) identifies language with speech and gesture, and would therefore be a central part of the study of the evolution of the sensorimotor component (SM) of FL.

Model (ii) identifies language with languages and, in this sense, the model is not particularly interested in FL as a biological object, nor, therefore, in the evolution of its components.

Model (iii) identifies language with communication: The studies inspired by Tomasello and Levinson are studies of the evolution of communication, not so much of FLB itself. Such work would thus constitute a part of the investigation of the evolution of the relation between the CI and SM components of FLB, while they simply ignore the computational dimension of language (FLN). Actually, in model (iii) language is a “cultural artefact” inserted into basic human communication, which is what would have evolved.

Model (iv) invokes the notion of a language-ready brain. This notion can be interpreted in two ways: Either the brain first developed, through evolution, those properties that make it capable of producing human languages (Chomsky’s point of view), or languages developed as complex cultural objects and then they served as an adaptive environment for the evolution of the language-ready brain from a “language-unready” brain (as in Deacon’s 1997 approach). In the first case, the

relevant parts of brain architecture and physiology (FLB) determine the distinctive properties of human language (and of human languages); in the second case, languages somehow externally developed this complexity and motivated the adaptations that would lead to the language-ready brain. In my opinion, the second version is highly implausible, although it is increasingly popular (as Wacewicz et al.'s report shows).

Actually, Arbib's version of the language-ready brain notion also equates language with languages, since his notion of a language-ready brain presupposes the existence of languages in the environment, as a brain-independent phenomenon:

What evolved (Evo) was a language-ready brain—not a brain with an innate mechanism encoding a universal grammar [...] but rather one enabling a child to acquire language (Devo), but only if raised in a milieu in which language is already present, something which, it is claimed, required tens of millennia of cultural evolution after the emergence of *Homo sapiens* (Socio).

(Arbib 2018, *apud* Wacewicz et al., 2020: 83)

Note that Arbib's qualification of what evolved (a language-ready brain and not "a brain with an innate mechanism encoding a universal grammar") is unnecessary, since, according to the HCF model of FLB (and according to the explicit model of Berwick and Chomsky 2016), there is no difference between, on the one hand, a brain programmed to develop a FLB and, on the other, a language-ready brain. A brain that includes that which makes only humans capable of learning and using the languages that they use (as Arbib points out, "only the human brain is language ready", Arbib 2012: ix, *apud* Wacewicz et al., 2020: 83) is already a brain that necessarily develops allowing the acquisition and use of human languages, that is, a brain with a FLB (including FLN). To put it more clearly: there is no difference (terminology aside) between a language-ready brain and a brain programmed to develop a FLB.

Thus, if the assumption that human languages precede the language-ready brain is not included, then the language-ready model is identical (except in detail) to Chomsky's. This is the case with the version of the language-ready brain proposed by Benítez-Burraco and Boeckx. In fact, the explanation of the internal structure and evolution of the FLB is very similar in both cases: According to Wacewicz et al. (2020), in the language-ready brain approach to language evolution set out by Benítez-Burraco and Boeckx, language (its "core combinatorial operation" [i.e. FLN]) is the result of genetic mutations that alter the anatomy and physiology of the brain, while "[o]ther components of language, particularly, some forms of phonology and pragmatics, are assumed to predate this human specific innovation" [i.e. SM and CI, the other older components of FLB according to HCF]. In this model of the language-ready brain:

[T]his freely combining merging ability is argued to be constrained via its interfacing with other cognitive systems and with the devices

involved in speech/gesture production. (Wacewicz et al., 2020: 85)

A characterization not unlike Chomsky's classic:

We take L [a language] to be a generative procedure that constructs pairs [...] that are interpreted at the articulatory-perceptual (A-P) and conceptual-intentional (C-I) interfaces. (1995: 219)

This language-ready brain model and Chomsky's model are also similar in terms of the idea that language (apparently its "core combinatorial operation") did not evolve for communication, but as a system of thought:

One interesting and distinctive consequence is a conception of language primarily as a tool for thinking (rather than for communicating), since our language-readiness would have initially emerged as a new, improved mechanism of conceptualising. (Wacewicz et al., 2020: 87)

This characterization is clearly reminiscent of the Chomskyan conception of FL and its evolution reflected in HCF, and developed in Chomsky (2007): "the earliest stage of language would have been just that: a language of thought, used internally" (Chomsky 2007: 13; see also Berwick & Chomsky 2011; 2016). As Chomsky has summarized more recently:

Investigation of the design of language gives good reason to take seriously a traditional conception of language as essentially an instrument of thought. Externalization then would be an ancillary process, its properties a reflex of the largely or completely independent sensorimotor system. (Chomsky 2016: 73)

Apart from the fact that Benítez-Burraco & Boeckx's approach to the evolution of the FLN (as presented in Wacewicz et al., 2020) is much more specific in biological and genetic aspects than that of Chomsky, it is evident that both models are variants of what is essentially the same theory of language.

Thus, the review of these four major research models of the evolution of language shows the usefulness of the HFC strategy of "fractionating FLB into several separate components, each of which might have different evolutionary histories" (Fitch et al. 2005: 205).

If it is not conceivable to give a coherent definition of language (and I agree), it is even less conceivable to speak of language evolution without further specification. The use of the expression *language evolution* without qualifying what *language* refers to implies, for example, that there is no clear distinction between the evolution of the organism that uses languages, and the historical change in the languages that this organism uses (as happens in some of the traditions mentioned above).

4. Conclusion

Wacewicz et al. (2020) argue that the absence of a restrictive definition of the object of study would be one of the causes of the progress that has been made in language evolution research, and they deny that the HCF initiative has had positive effects in that direction. This conclusion is based on the claim of an unnoticed inconsistency in the definition of FLN, but such an inconsistency does not exist.

Of course, it is possible that the absence of a hegemonic “top-down” definition of language has opened the door to more pluralistic and interdisciplinary research activity. But it does not seem (in light of the synthesis presented in Wacewicz et al, 2020) that we have sufficient perspective today to affirm that remarkable progress has been made over the last 10 years in the understanding of how the human faculty of language evolved. I tend to agree with Lewontin in that “[w]e know essentially nothing about the evolution of our cognitive capabilities, and there is a strong possibility that we will never know much about it” (Lewontin 1998: 109), although I do not fully share his attitude towards the achievements in the field of language evolution research:

Reconstructions of the evolutionary history and the causal mechanisms of the acquisition of linguistic competence or numerical ability are nothing more than a mixture of pure speculation and inventive stories. (Lewontin 1998: 111)

I have no doubt, in view of the valuable synthesis presented in Wacewicz et al. (2020), that the field of language evolution research is in good health and is bringing together important intellectual and economic efforts. But I do have doubts that the lack of a more uniform and specific definition of what evolved when language evolved is a positive aspect here. I see no reason to affirm that the abandoning of an explicit proposal of the division of labour and the avoidance of misunderstandings, as in the FLB/FLN terminological proposal, is an advantage; rather, it might well be the opposite.

Acknowledgements

The author’s research is supported by the Spanish State Research Agency (AEI) & FEDER (EU) grant FFI2017-82460-P, and a Government of Aragon grant to the Psylex (Language and Cognition) research group.

References

- Berwick, Robert C. & Noam Chomsky. 2011. The biolinguistic program: The current state of its development. In Anna Maria Di Sciullo & Cedric Boeckx (eds.), *The Biolinguistic Enterprise. New Perspectives on the Evolution and Nature of the Human Language Faculty*, 19–41. Oxford: Oxford University Press.
- Berwick, Robert C. & Noam Chomsky. 2016. *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans Martin Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Semantics*, 1–30. Berlin: Mouton de Gruyter.
- Chomsky, Noam. 2016. *What Kind of Creatures Are We?* New York: Columbia University Press.
- Deacon, Terrence W. 1997. *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: Norton.
- Fitch, Tecumseh W., Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: Clarifications and implications. *Cognition* 97(2), 179–210.
- Hauser, Marc, Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Lewontin, Richard. 1998. The evolution of cognition: Questions we will never answer. In Don Scarborough & Saul Sternberg (eds.), *Methods, Models, and Conceptual Issues: An Invitation to Cognitive Science*, 108–132. Cambridge, MA: MIT Press.
- Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 95(2), 201–236.
- Waciewicz, Sławomir, Przemysław Żywiczyński, Stefan Hartmann, Michael Pleyer, & Antonio Benítez-Burraco. 2020. *Language in language evolution research: In defense of a pluralistic view*. *Biolinguistics* 14.SI, 59–101.

Biolinguistics

Volume 14.SI

2020

TABLE OF CONTENTS

- 051 Why Don't Languages Grammaticalize [\pm poisonous]? Evelina Leivada
Universitat Rovira i Virgili
Lluís Barceló-Coblijn
Universitat de les Illes Balears
- 059 *Language in Language Evolution Research: In Defense of a Pluralistic View* Sławomir Wacewicz
Nicolaus Copernicus University of Torun
Przemysław Żywicznyński
Nicolaus Copernicus University of Torun
Stefan Hartmann
University of Düsseldorf
Michael Pleyer
University of Koblenz-Landau
Antonio Benítez-Burraco
University of Seville
- 102 The Phonological Latching Network Joe Stephen Bratsvedal Collins
UiT – The Arctic University of Norway, International School for Advanced Studies (SISSA) in Trieste, & Norwegian University of Science and Technology in Trondheim
- ★ BRIEFS ★ 130 Humans Discriminate Individual Zebra Finches by Their Song Sabrina Schalz
Middlesex University London
Thomas E. Dickins
Middlesex University London
- ★ FORUM ★ 145 In Defence of FLB/FLN: A Reply to Wacewicz et al. (2020) José-Luis Mendívil-Giró
University of Zaragoza

