The evolution of combinatorial structure in language

Zuidema, W.; de Boer, B.

Published in:
Current Opinion in Behavioral Sciences

DOI:
10.1016/j.cobeha.2018.04.011

Citation for published version (APA):
Zuidema, W., & de Boer, B. (2018). The evolution of combinatorial structure in language. *Current Opinion in Behavioral Sciences*, *21*, 138-144. https://doi.org/10.1016/j.cobeha.2018.04.011

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Behavioral Sciences

The evolution of combinatorial structure in language

Willem Zuidema[1] and Bart de Boer[2]

Human language shows combinatoriality in its phonology (both in speech and in sign language) and its grammar, while both types appear to be absent in the communication systems of our closest evolutionary relatives. In this article, we observe that productive combinatoriality is difficult to evolve, because it requires multiple components to be put in place simultaneously for it to function. To understand how it nevertheless evolved in human language, we focus on combinatoriality in phonology, for which most evidence is available. We discuss findings and theories from three domains: linguistics (descriptive, experimental and corpus linguistics), comparative biology (including some fossil indicators) and (computer) models. We tentatively conclude that many of the biological prerequisites for combinatorial phonology and compositional semantics are shared with other animals, but that a uniquely human pressure for large vocabularies and uniquely human processes of cultural evolution are key in understanding the origins of combinatoriality in language.

Addresses
[1] University of Amsterdam, the Netherlands
[2] VUB – Vrije Universiteit Brussel, Belgium

Corresponding author: Zuidema, Willem (zuidema@uva.nl)

Current Opinion in Behavioral Sciences 2018, **21**:138–144

This review comes from a themed issue on **The evolution of language**

Edited by **Christopher Petkov** and **William Marslen-Wilson**

For a complete overview see the Issue and the Editorial

Available online 1st June 2018

https://doi.org/10.1016/j.cobeha.2018.04.011

Introduction

A celebrated property of human language is its productive combinatoriality: we combine vowels and consonants into syllables, syllables into words, words into sentences. Combinatoriality makes language open-ended: we can always create new words from the speech sounds of our language when a new concept needs a name, and we can communicate about uncountably many complex thoughts using novel combinations of words. In this article, we consider both the *property of language* of using a limited set of building blocks to create a much larger, perhaps even unbounded, set of utterances and the *property of humans* of being able to deal with signals that are structured in this way. When we discuss evolution, this also has two sides: the (cultural) evolution of combinatorial structure in languages and the (biological) evolution of mechanisms to deal with combinatorial structure.

The term combinatorial structure may refer to combinations of speech sounds (combinatorial phonology), combinations of signs in sign languages (which could be called combinatorial cherology) or combinations of meaningful morphemes or words (*compositional*[1] *semantics*) [1]. We will focus mostly on combinatorial phonology, but very similar points can be made about other combinatorial systems, although less evidence is available informing us about those systems. In human spoken language, the building blocks are generally identified with phonemes: minimal components of speech that are produced in sequence and where, when one such component is replaced by another, the meaning of a word changes [2,3]. Thus, in the words ‘pat’ and ‘bat’, /p/ and /b/ are phonemes, because replacing one for the other changes the meaning of the word. Using such pairs of words (‘minimal pairs’) we can demonstrate that /æ/ (which represents the pronunciation of the ‘a’ in these words) and the /t/ are also phonemes. In signed language, the building blocks are generally identified with handshapes, hand movements and location [4]. In both cases, the building blocks can only be combined according to well-defined, language-specific rules.

Although linguists agree that language has combinatorial structure, there is discussion about what the actual building blocks are: the building blocks in combinatorial phonology could be phonemes, features (individual articulatory/acoustic events that are combined to produce phonemes), syllables or parts of syllables (onsets, nuclei and codas — the final consonants — for instance), or perhaps all of them depending on the context and the language (see the introduction in [5]). Similarly, the building blocks in compositional semantics could be morphemes and words, or constructions.

In addition, different streams can be combined in parallel [6]: in spoken language, intonation and tone are combined with the sequence of phonemes [7,8], while in sign language body posture, head posture and facial expressions are combined with manual signs [9]. Combinatorial structure does thus not necessarily depend on phonemes in sequence, and may combine several streams in parallel. This observation is especially relevant when studying

[1] We will use the term ‘combinatorial’ for any system where parts are combined into larger wholes, and the term ‘compositional’ specifically for the combination of *meaning-carrying* parts, where the meaning of the whole depends on the meaning of the parts.

potential precursors of combinatorial structure and potential scenarios through which it may have evolved.

Where human language is thus both combinatorial and semantic, very few non-human animal communication systems have both these properties (but see [10•,11]). In birds, cetaceans [12] and gibbons [13,14] we find vocalizations with 'bare phonology' [15]: song elements are productively combined into songs, but elements nor songs are semantic (or 'referential'). It appears, on the other hand, that great ape vocalizations, which may be semantic, do not use combinatorial structure [16].

The question therefore arises how the ability to use combinatorial structure has evolved in our species (we will focus here on biological evolution and on cultural evolution in as far as it is necessary to understand biological evolution; we will not focus on using evolution as a theoretical framework to understand language change [17]). This question has many subquestions: what causes combinatorial structure? Which cognitive mechanisms are involved? Did these undergo selection due to speech or not? Did our ancestors have similar mechanisms? And what (combination of) functional factors was involved? What is the role of communication, learnability and the modality (speech or sign)? When did combinatorial structure evolve, and over how much time? And how did the evolution happen? What selective pressures were involved? What were the precursors? And did syntax evolve before phonology or the other way around? This article reviews evidence pertinent to these questions, focusing on three sources of evidence: linguistics (descriptive, experimental and corpus linguistics), comparative biology (including some fossil indicators) and (computer) models. But before reviewing these categories of evidence, we first briefly consider why combinatoriality poses particular challenges for evolutionary theories and why it is rare in nature.

Combinatoriality and evolution

The trick of productively combining discrete elements from a finite repertoire into a large number of combinations is rare in Nature — it is a trick that phonology and grammar only seem to share with music, bird song, cetacean song, the genetic code, and, in primitive form, perhaps in the vocalizations of a handful of non-human primates [18]. Productive combinatoriality is difficult to evolve, because it requires multiple components to be put in place simultaneously for it to function. In both communication systems and systems that mainly serve as displays, that is, to impress, there needs to be, firstly, a repertoire of basic elements shared by sender and receiver, and secondly, a mechanism to combine those elements into larger combinations in the sender (*synthesis*), for the system to be productively combinatorial. In a system such as language, we additionally need, thirdly, the mechanisms to break down combinations into their component parts in the receiver (*analysis*).

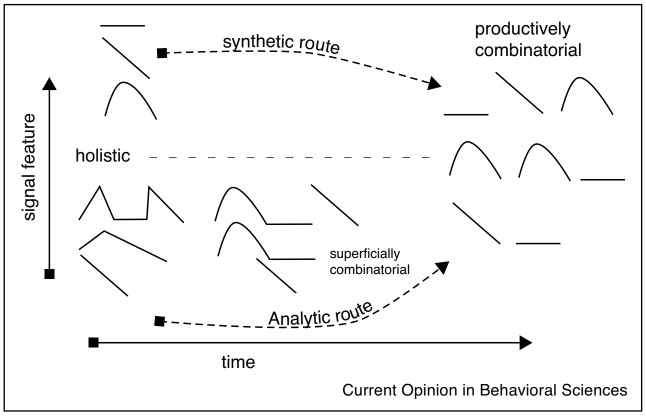
Biological evolution proceeds one mutation at a time, with each mutation starting as a unique variant and having to spread in the population by conveying a fitness advantage to the individuals that carry it. The challenge for theories of the evolution of combinatoriality is therefore to explain how components (1), (2) or (3) could evolve without the other components being in place already.

We currently do not know what the neural basis is for combinatorial phonology, combinatorial cherology and compositional semantics, making it difficult to assess how likely it is that components (1), (2) or (3) are distinct systems that needed to evolve one after the other, or side-effects of the same underlying biological innovation. Most theories of the evolution of combinatoriality in language simply assume, implicitly or explicitly, that the ability for (2) or (3) are side-effects of features of the human brain that evolved for other reasons; for instance, the ability to interpret combinatorial signals (3) could be based on preexisting cognitive mechanisms to process information about the environment. For a completely satisfactory account of the evolution of productive combinatoriality this assumption would need to be supported — making empirical work on human perceptual biases independent of language and speech [19•,20••,21•] and unravelling the neural basis of combinatoriality [22••,23••] key areas of research on language evolution.

Without solving this challenge, Hurford [24] distinguishes two classes of (non-mutually exclusive) scenarios: the analytic route versus the synthetic route to combinatoriality. In the analytic route, *superficial* combinatorial structures arises by chance or by some other process; mechanisms to make productive use of combinatorial structure can then invade in the population while maintaining interpretability [25]. In the synthetic route, the later building blocks of combinatorial structure are assumed to initially have been used as independent holistic signals; mechanisms to productively combine these building blocks evolve only later. In Figure 1 we sketch Hurford's two routes.

Both the synthetic and the analytic route have been studied in computer models of the evolution of combinatoriality. Nowak and Krakauer [26] derive mathematically an 'error limit' for holistic signaling, and show that combinatoriality can help overcome that limit. Their results may be seen as support for the synthetic route. Zuidema and de Boer [25], on the other hand, lend support to the analytic route, by demonstrating that a large holistic vocabulary in a restricted signal space ('crowding'), leads to signal overlap and thus *superficial combinatoriality*. Little *et al.* [20••] in turn question the

Figure 1



Different scenarios for the evolution of a combinatorial signaling system. Both scenarios start from a 'holistic' communication system without combination (left of the figure), and end with a productively combinatorial system (right). In the 'synthetic route' (top), preexisting signals at some point start being combined with each other. In the 'analytic route' (bottom), holistic signals are assumed to first evolve into signals that are superficially combinatorial: different signals have overlapping parts, but the productive use of this overlap only comes later.

relevance of crowding, and report evidence that humans have a strong tendency to detect structure in signals even before the signal space gets crowded. This work still supports a primarily analytic scenario, but suggests an alternative way to arrive at superficial combinatoriality, that is empirically supported but not yet formalized in a computational model.

Linguistic theories and linguistic evidence

Hockett [27] was the first to write about combinatorial structure in an evolutionary context. He observed that human languages not only combine meaningless (acoustic) building blocks into meaningful utterances (what we call 'combinatorial phonology'), but then also combine these utterances (morphemes and words) into longer meaningful utterances (phrases and sentences; 'compositional semantics'). Both types of combination follow learned rules. As there are two levels of combination involved, he called this the *duality of patterning*. Moreover, he observed that this was unique to human language (leaving open the possibility that combination on one level occurs in animal communication). He proposed that it evolved in human language in order to accommodate a large range of signals.

Other authors have tried to flesh out these ideas by proposing concrete scenarios by which duality of patterning or combinatorial structure evolved. The frame-content theory [28,29] proposes that combinatorial structure in speech makes use of pre-existing rhythmic behaviors of the jaw involved in eating, sucking and breathing. These give us the basic syllables (frames), on which more refined articulatory gestures (content) are superimposed. This, according to [30] also explains differences in rhythmic structures between sign language and spoken language; although both use combinatorial structure, it is proposed that the precise properties do not derive from cognitive constraints, but from physical ones related to the modality.

The gestural origins theory of speech [18,31] proposes that the building blocks derive from articulatory gestures (not to be confused with visual co-speech gestures). The articulators that produce these gestures can be considered coupled oscillators, and depending on the phases with which these oscillators are coupled, different patterns can be produced. Moreover, because the jaw is the most massive oscillator, it would dominate the coupling, and therefore produce something very similar to the frame-content theory. Interestingly in this context it has also been proposed [31] that originally, articulatory-vocal gestures formed the basis of phonology, while manual gestures formed the basis of syntax. Both the vocalic gestural origins and the frame-content theory derive support from patterns with which consonants and vowels tend to co-occur in words, and in how such combinations appear in infant vocalizations.

A different route to duality of patterning has been proposed by [32], where it is proposed that synonymy avoidance creates pressure for larger lexicons, which in a co-evolutionary process leads to cognitive adaptations to combinatorial structure. Accidental combinations of words then lead to compositional structure. Because compositional structure follows, and is based on combinatorial structure, it is argued that compositional semantics is based on cognitive mechanisms that originally evolved for combinatorial phonology and therefore follows similar patterns. It should be noted that this view is not necessarily contradictory with the other scenarios, focusing on functional pressures rather than actual precursors and mechanisms.

Observations of emerging and existing languages may also provide information about how combinatorial speech may have evolved. Linguistic fossils — aspects of language that do not quite follow grammatical rules [33] — show that learned, linguistic utterances do not necessarily need to follow phonological rules (such as 'psst' and 'shhh', which would not be acceptable English words) or be combinatorial (such as tongue clacking of disapproval, which does not use English phonemes — [! !] in IPA notation).

Observation of emerging spoken languages, such as pidgins or jargons do not provide us with much useful information about the emergence of combinatorial phonology, as they make use of speech sounds of the languages on which they are based (although in simplified form, following universal tendencies [34,35]) but they did inspire theories on the evolution of syntax [36].

Emerging sign language do provide us with such information, because they are often not based on an existing language, and are invented from scratch. The classic example is Nicaraguan Sign Language [37,38], although the focus of research on this language has been on syntax and morphology. Other emerging sign languages, such as Central Taurus Sign Language [39] and Al-Sayyid Bedouin Sign Language (ABSL) [40,41] have been studied for emergence of phonology. Especially in ABSL, it is clear that an emerging sign language can exist without combinatorial phonology (cherology) [41] but with compositional semantics. Initially, there appears to be great individual variability in signs, and no identifiable building blocks. However, combinatorial structure does appear to emerge gradually in later generations of speakers.

Because emergence of language is a very rare phenomenon and difficult to study 'in the wild', emergence of language is now often studied in laboratory settings [42,43]. Many of these studies do not focus on combinatorial structure, but an increasing number does. In order to prevent interference from native language, these studies use gestures produced by non-signers [44], or different artificial signaling devices, such as drawing pads [45,46], slide whistles [47,48] and infra-red sensors [49]. These experiments investigate how duality of patterning emerges [46], whether combinatorial structure is due to cognitive mechanisms or functional pressures [48], what the influence of iconicity is on emergence of structure [50], and what the effect of modality is on these processes [44,50,51].

These studies show there is a complicated interaction between modality, communicative setting and cognitive mechanisms. Iconicity appears to be a viable strategy when transparent mappings between signals and meanings are possible, and in this case emergence of combinatorial structure is delayed. However, the human tendency to find and generalize patterns causes combinatorial structure to emerge eventually, especially in cases where there is repeated interaction between individuals, who simplify and abbreviate signals they are familiar with.

Comparative biological evidence

Humans have some obvious modifications of the vocal tract compared to other great apes, and these modifications have been discussed in the context of fossil evidence [52,53] leading to the tentative conclusion that Neanderthals already had some kind of ability for complex vocalization [54]. However, this evidence does not tell us much about evolution of combinatorial structure. Conversely, there is emergent evidence that shows that great apes do have rudimentary abilities that could be precursors.

It has been argued that an ape-like vocal tract is not able to produce a sufficient variety of speech sounds for language [55], but recent work has shown that monkey vocal tracts — and by implication the vocal tract of the last common ancestor with great apes — have the capacity to produce enough different speech sounds for language [56••,57]. Moreover, although it has been argued that great apes lack voluntary control over the vocal folds [15], recent evidence shows that gorillas [58] and orangutans [59] can learn to control their vocal folds to some extent. Evidence for chimpanzee vocal adaptation [60,61] points in the same direction. This indicates that the last common ancestor probably already had rudimentary abilities for producing a range of learned, controlled vocalizations.

The question remains whether these vocalizations had any structure to them, and whether apes are able to deal with this. It has been shown that lip smacks in macaques [62] and orangutans [63] show the same temporal structure as syllables. Also, it has been argued that orangutans combine consonant-like sounds with vowel-like sounds [64]. However, these results are too preliminary at the moment to warrant strong conclusions.

Biological evidence has also been brought to bear on the more abstract question of how combinatorial structure and duality of patterning can evolve in principle. It has been argued that no other animal shows phonological combinatorial structure ([16], but see [65••]) and that therefore syntax may have evolved before phonology. However, this rests on the assumption that combinatorial structure is based on distinguishing meaning, and this disqualifies 'bare phonology' [15] such as found in birds, cetaceans [12] and gibbons [13,14]. It has even been argued that Japanese great tits have duality of patterning [66].

One thing that does appear from all this work is that systems with many different calls that are composed of a limited number of building blocks are most commonly used for display, rather than to convey many different meanings. This has led to the proposal that something similar has happened in evolution of human language, and that combinatorial speech was initially use for display — the musical origins hypothesis [15,67].

Conclusion

From the emerging biological evidence, it is clear that many of the basic behaviors needed to produce combinatorial structure are already present in apes — and by homology our last common ancestor with them — except for the ability to learn a large set of spoken (or signed) utterances. There are different scenarios about how these pre-existing behaviors (articulatory gestures, oscillatory behaviors, simple acoustic imitations) form the basis of the combinatorial structure of speech. These scenarios are not mutually exclusive, but on the contrary overlap to such an extent that it is hard to choose between them empirically.

Although bare phonology, that is combinatorial signals without associated different meanings, is found in many other animals, it is not clear whether there ever was a stage of bare phonology before speech. It seems equally likely that pressure on increasing the number of signals came from increased need to communicate different meanings. After all, apes can already learn reasonably large communicative lexicons if they are trained (e.g. [68]) so presumably the last common ancestor also had the ability to use signals communicatively.

It has been proposed that the need for an increasing number of signals caused evolution of combinatorial structure and our ability to deal with it [18,27]. There are even mathematical and computer models that simulate this [22••,25,26]. However, experiments show that modern humans have a strong tendency to see and generalize structure before the signal space gets crowded [19•,20••,21•,48] so it is conceivable that combinatorial structure actually is based on much older cognitive mechanisms to detect structure in the environment. It is of course possible that such mechanisms have been fine-tuned through selection for speech. Through cultural transmission, languages will then have evolved (culturally) to show more and more combinatorial structure, as seems to be happening in emerging sign languages [41] while (at least in spoken languages) the building blocks become more and more distinct [69,70,71•]. Thus, the ability to use combinatorial structure and combinatorial structure itself could have co-evolved gradually.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. De Boer B, Sandler W, Kirby S: **New perspectives on duality of patterning: Introduction to the special issue**. *Lang Cogn.* 2012, **4**:251-259.

2. Odden D: *Introducing Phonology*. Cambridge University Press; 2013.

3. Ladefoged P, Maddieson I: *The Sounds of the World's Languages*. Blackwell; 1996.

4. Sandler W: **The phonological organization of sign languages**. *Lang Linguist Compass* 2012, **6**:162-182.

5. Goldinger SD, Azuma T: **Puzzle-solving science: The quixotic quest for units in speech perception**. *J Phon.* 2003, **31**:305-320.

6. Kremers J: **The syntax of simultaneity**. *Lingua* 2012, **122**:979-1003.

7. Yip M: *Tone*. Cambridge University Press; 2002.

8. Cruttenden A: *Intonation*. Cambridge University Press; 1997.

9. Sandler W: **Prosody and syntax in sign languages**. *Trans Philos Soc.* 2010, **108**:298-328.

10. • Schlenker P, Chemla E, Schel AM, Fuller J, Gautier J-P, Kuhn J, Veselinović D, Arnold K, Cäsar C, Keenan S, Lemasson A, Ouattara K, Ryder R, Zuberbühler K: **Formal monkey linguistics**. *Theor Linguist* 2016, **42**:1-90.

This collaboration between linguists and biologists reviews a large body of work on monkey communication systems, and argues that they can be analyzed using concepts and techniques from various linguistic subfields, including semantics, syntax, morphology, pragmatics and phonology. The paper presents a great overview of the complexities of monkey vocalizations, and calls attention to the role of 'pragmatics' in the interpretation of monkey calls, but is cautious in claiming relevant parallels with human language. The main goal of the paper is to argue for a systematic, comparative study of non-human communication and human language.

11. Schlenker P, Chemla E, Zuberbühler K: **What do monkey calls mean?** *Trends Cogn Sci* 2016, **20**:894-904.

12. Payne RS, McVay S: **Songs of humpback whales**. *Science* 1971, **173**:585-597.

13. Mitani JC, Marler P: **A phonological analysis of male gibbon singing behavior**. *Behaviour* 1989, **109**:20-45.

14. Geissmann T: **Duet-splitting and the evolution of gibbon songs**. *Biol Rev Camb Philos Soc.* 2002, **77**:57-76.

15. Fitch WT: *The Evolution of Language*. Cambridge University Press; 2010. (Chapters 9, 14).

16. Collier K, Bickel B, van Schaik CP, Manser MB, Townsend SW: *Language Evolution: Syntax Before Phonology?* The Royal Society; 201420140263.

17. Blevins J: *Evolutionary Phonology*. Cambridge University Press; 2004.

18. Studdert-Kennedy M: **How did language go discrete?** In *Language Origins: Perspectives on Evolution.* Edited by Tallermann M. Oxford University Press; 2005:48-67.

19. • Eryilmaz K, Little H: **Using Leap Motion to investigate the emergence of structure in speech and language**. *Behav Res Methods* 2017, **49**:1748-1768 http://dx.doi.org/10.3758/s13428-016-0818-x.

Paper presenting an innovative method — based on converting hand movements to sound — for studying the emergence of combinatorial, discrete structure in a continuous domain.

20. •• Little H, Rasilo H, van der Ham S, Eryilmaz K: **Empirical approaches for investigating the origins of structure in speech**. *Interact Stud* 2017, **18**:332-354 http://dx.doi.org/10.1075/is.18.3.03lit.

This paper reviews the both experimental and computational modelling work investigating the emergence of structure in speech, both addressing how inventories of individual speech sounds emerge and how individual speech sounds are reused in combinatorial structure.

21. • Little H, Eryilmaz K, de Boer B: **Signal dimensionality and the emergence of combinatorial structure**. *Cognition* 2017, **168**:1-15 http://dx.doi.org/10.1016/j.cognition.2017.06.011.

Little*et al.* present an artificial signaling experiment to test the conditions under which combinatorial structure will emerge. In the experiments, participant learn to use hand movements to signify a range of meanings. Hand movements are transformed to sound using the approach from Ref. [19•]. The experimenters manipulate the dimensionality of the 'meaning space' and 'signal space', and, importantly, report that the more similar meaning and signal spaces are, the more likely it is that iconic rather than combinatorial signal systems emerge.

22. •• Havrylov S, Titov I: **Emergence of language with multi-agent games: learning to communicate with sequence of symbols**. In *Proceedings Neural Information Processing Systems (NIPS 2017)*. 2018.

Havrylov and Titov revisit the question of how compositional semantics can culturally evolve in a communication system, using modern deep learning techniques. Using a clever trick to speed up the optimization (where information about gradients is passed on between speakers and hearers), they show that communication systems optimized for communicative success between agents spontaneously exhibit compositional structure.

23. •• Hupkes D, Veldhoen S, Zuidema W: **Visualisation and 'diagnostic classifiers' reveal how recurrent and recursive neural networks process hierarchical structure**. *J Artif Intell Res* 2018, **61**:907-926.

The authors show that a modern deep neural network architecture, the Gated Recurrent Unit (GRU), can learn an artificial language with compositional semantics (the language of arithmetic) and generalize to longer expressions than seen at training. They, moreover, develop a methodology (diagnostic classification) to interpret the strategies that the networks acquire. Applying diagnostic classification to GRU's reveals how a neural system (without the usual 'symbolic' variables and rules) may be able to implement compositional semantics.

24. Hurford JR: *The Origins of Grammar – Language in the Light of Evolution*. Oxford University Press; 2012.

25. Zuidema W, de Boer B: **The evolution of combinatorial phonology**. *J Phon.* 2009, **37**:125-144.

26. Nowak MA, Krakauer D, Dress A: **An error limit for the evolution of language**. *Proc R Soc Lond.* 1999, **266**:2131-2136.

27. Hockett C: **The origin of speech**. *Sci Am.* 1960, **203**:88-111.

28. MacNeilage PF, Davis BL: **On the origin of internal structure of word forms**. *Science* 2000, **288**:527-531.

29. MacNeilage PF: **The frame/content theory of evolution of speech production**. *Behav Brain Sci.* 1998, **21**:499-511.

30. MacNeilage PF: *The Origin of Speech*. Oxford University Press; 2008.

31. Goldstein L, Byrd D, Saltzman EL: **The role of vocal tract gestural action units in understanding the evolution of phonology**. In *Action to Language via the Mirror Neuron System.* Edited by Arbib MA. Cambridge University Press; 2006:215-249.

32. Carstairs-McCarthy A: *The Origins of Complex Language: An Inquiry in the Evolutionary Beginnings of Sentences, Syllables, and Truth*. Oxford University Press; 1999.

33. Jackendoff R: *Foundations of Language*. Oxford University Press; 2002.

34. Singh R, Muysken P: **Wanted: a debate in pidgin/creole phonology**. *J Pidgin Creole Lang.* 1995, **10**:157-169.

35. Klein TB: **Creole phonology typology: Phoneme inventory size, vowel quality distinctions and stop consonant series**. In *The structure of creole words: segmental, syllabic and morphological aspects.* Edited by Bhatt P, Plag I. Walter de Gruyter; 2006:3-21.

36. Bickerton D: **The language bioprogram hypothesis**. *Behav Brain Sci.* 1984, **7**:173-222.

37. Polich L: **The emergence of the deaf community in Nicaragua**. *Gallaudet* 2005.

38. Senghas A, Kita S, Özyürek A: **Children creating core properties of language: evidence from an emerging sign language in Nicaragua**. *Science* 2004, **305**:1779-1782.

39. Caselli N, Ergin R, Jackendoff R, Cohen-Goldberg A: *The emergence of phonological structure in Central Taurus Sign Language*. 2014.

40. Sandler W, Meir I, Padden C, Aronoff M: **The emergence of grammar: systematic structure in a new language**. *Proc Natl Acad Sci USA.* 2005, **102**:2661-2665.

41. Sandler W, Aronoff M, Meir I, Padden C: **The gradual emergence of phonological form in a new language**. *Nat Lang Linguist Theory* 2011, **29**:503-543.

42. Scott-Phillips TC, Kirby S: **Language evolution in the laboratory**. *Trends Cogn Sci.* 2010, **14**:411-417.

43. Galantucci B: **Experimental semiotics: a new approach for studying communication as a form of joint action**. *Top Cogn Sci.* 2009, **1**:393-410.

44. Namboodiripad S, Lenzen D, Lepic R, Verhoef T: **Measuring conventionalization in the manual modality**. *J Lang Evol.* 2016, **1**:109-118.

45. Roberts G, Galantucci B: **The emergence of duality of patterning: insights from the laboratory**. *Lang Cogn.* 2012, **4**:297-318.

46. Del Giudice A: **The emergence of duality of patterning through iterated learning: precursors to phonology in a visual lexicon**. *Lang Cogn.* 2012, **4**:381-418.

47. Verhoef T: **The origins of duality of patterning in artificial whistled languages**. *Lang Cogn.* 2012, **4**:357-380.

48. Verhoef T, Kirby S, de Boer B: **Emergence of combinatorial structure and economy through iterated learning**. *J Phon.* 2014, **43**:57-68.

49. Eryilmaz K, Little H: **Using leap motion to investigate the emergence of structure in speech and language [Internet]**. *Behav Res Methods* 2016 http://dx.doi.org/10.3758/s13428-016-0818-x.

50. Roberts G, Lewandowski J, Galantucci B: **How communication changes when we cannot mime the world: experimental evidence for the effect of iconicity on combinatoriality**. *Cognition* 2015, **141**:52-66.

51. Little H, Eryilmaz K, de Boer B: **Signal dimensionality and the emergence of combinatorial structure**. *Cognition* 2017, **168**:1-15.

52. Fitch WT: **Fossil cues to the evolution of speech**. In *The Cradle of Language.* Edited by Botha R, Knight C. Oxford University Press; 2009:112-134.

53. de Boer B: **Evolution of speech and evolution of language**. *Psychon Bull Rev.* 2017, **24**:158-162.

54. Dediu D, Levinson SC: **On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences**. *Front Psychol.* 2013, **4**:1-17.

55. Lieberman PH, Klatt DH, Wilson WH: **Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates**. *Science* 1969, **164**:1185-1187.

56. •• Fitch WT, de Boer B, Mathur N, Ghazanfar AA: **Monkey vocal tracts are speech-ready**. *Sci Adv.* 2016, **2**.

Fitch*et al.* study the shapes the vocal tract of Macaques can take, taking into account feeding movements, facial displays and vocalization, using X-ray videos. Using a detailed articulatory model, they show that, contrary to earlier claims (such as Ref. [55]), the range of shapes offers an adequate range of speech sounds to support language-like communication. That is, the monkey vocal tract is 'speech ready', and is unlikely to have been the constraining factor in the evolution of speech and language. Refs. [55,56••] bookend an almost 50 year long debate about the importance of anatomy in determining vocal abilities in ancestral (fossil) species. Anatomy appears to be less important than was long thought.••

57. Boë L-J, Berthommier F, Legou T, Captier G, Kemp C, Sawallis TR, Becker Y, Rey A, Fagot J: **Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests pre-hominin speech precursors**. *PLOS ONE* 2017, **12**e0169321.

58. de Boer B, Perlman M: **Physical mechanisms may be as important as brain mechanisms in evolution of speech**. *Behav Brain Sci.* 2014:37.

59. Lameira A, Hardus M, Mielke A, Wich S, Shumaker R: **Vocal fold control beyond the species-specific repertoire in an orang-utan**. *Sci. Rep.* 2016, **6**:1-10.

60. Crockford C, Herbinger I, Vigilant L, Boesch C: **Wild chimpanzees produce group-specific calls: a case for vocal learning?** *Ethology* 2004, **110**:221-243.

61. Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, West V, Slocombe KE: **Vocal learning in the functionally referential food grunts of chimpanzees**. *Curr Biol.* 2015, **25**:495-499.

62. Ghazanfar AA, Takahashi DY, Mathur N, Fitch WT: **Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics**. *Curr Biol.* 2012, **22**:1176-1182.

63. Lameira AR, Hardus ME, Bartlett AM, Shumaker RW, Wich SA, Menken SB: **Speech-like rhythm in a voiced and voiceless orangutan call**. *PLOS ONE* 2015, **10**e116136.

64. Lameira AR, Vicente R, Alexandre A, Campbell-Smith G, Knott C, Wich S, Hardus ME: **Proto-consonants were information-dense via identical bioacoustic tags to proto-vowels**. *Nat Hum Behav*. 2017, **1**:44.

65. Engesser S, Ridley A, Townsend SW: **Meaningful call ••combinations and compositional processing in the southern pied babbler**. *Proc Natl Acad Sci U S A* 2016, **113** http://dx.doi.org/10.1073/pnas.1600970113.

This paper reports the most compelling evidence to date on a rudimentary form of compositional semantics in the communication system of a non-primate: the southern pied babbler — a social bird species. Importantly, the authors report playback experiment using single elements and novel combinations to make plausible that the bird indeed process these combinatorially, rather than by holistically storing every possible combination.

66. Suzuki TN, Wheatcroft D, Griesser M: **Experimental evidence for compositional syntax in bird calls**. *Nat Commun.* 2016, **7**:10986.

67. Mithen S: *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. Harvard University Press; 2007.

68. Patterson FG, Cohn RH: **Language acquisition by a lowland gorilla: Koko's first ten years of vocabulary development**. *Word* 1990, **41**:97-143.

69. Liljencrants J, Lindblom B: **Numerical simulations of vowel quality systems**. *Language* 1972, **48**:839-862.

70. de Boer B: **Self organization in vowel systems**. *J Phon.* 2000, **28**:441-465.

71. Wedel A, Fatkullin I: **Category competition as a driver of •category contrast**. *J Lang Evol* 2017, **2**:77-93 http://dx.doi.org/10.1093/jole/lzx009.

Wedel*et al.* study how categories emerge in perceptual-motor loops such as those between the acoustic perception and articulation mechanisms in vocal imitation.