

Evolution of Speech: Anatomy and Control

De Boer, Bart

Published in:
Journal of Speech, Language, and Hearing research

DOI:
[10.1044/2019_JSLHR-S-CSMC7-18-0293](https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0293)

Publication date:
2019

License:
Unspecified

Document Version:
Accepted author manuscript

[Link to publication](#)

Citation for published version (APA):
De Boer, B. (2019). Evolution of Speech: Anatomy and Control. *Journal of Speech, Language, and Hearing research*, 62(8S), 2932-2945. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0293

Copyright

No part of this publication may be reproduced or transmitted in any form, without the prior written permission of the author(s) or other rights holders to whom publication rights have been transferred, unless permitted by a license attached to the publication (a Creative Commons license or other), or unless exceptions to copyright law apply.

Take down policy

If you believe that this document infringes your copyright or other rights, please contact openaccess@vub.be, with details of the nature of the infringement. We will investigate the claim and if justified, we will take the appropriate steps.

Evolution of Speech: Anatomy and Control

Bart de Boer

AI-lab

Vrije Universiteit Brussel

Pleinlaan 2

1050 Brussel

Belgium

bart@ai.vub.ac.be

Abstract

This paper critically reviews work on the evolution of speech in the context of motor control. It presents a brief introduction of the field of language evolution, and argues why taking the evolutionary perspective is useful. It then proceeds to review different methods of studying evolutionary questions: comparative research, experimental and observational research, and computer and mathematical modeling.

On the basis of comparative analysis of related species (specifically, other great apes) and on the basis of theoretical results, the paper argues that adaptations for language and speech must have evolved gradually and that it is likely that speech motor control is one of the key aspects that has undergone selection related to speech. It predicts that this means that it must be possible to find empirical evidence for this, but that such research is only in its infancy at the present moment.

Introduction

Human speech is a very complex behavior, and it would seem that humans must have evolved adaptations for producing, perceiving and learning it. However, as with many questions related to language, it turns out that answering the questions of what has evolved for speech, when it evolved and for what reasons it evolved are not easy to answer. This article aims to illustrate how questions about the evolution of speech can be investigated from an evolutionary perspective, and focuses in particular on the use of computer models. It does not present an exhaustive review of the growing body of work in this field, but aims to put the questions and methods into perspective and show what the separate strands of research indicate when considered together. As it turns out, adaptations to control over speech may have been one of the crucial evolutionary adaptations that have led to the modern human ability for language. But before delving into the research on the evolution of speech, it is useful to briefly review why the evolutionary point-of-view is useful, how evolution can be investigated and how computer models can be of particular relevance in this field.

Why an evolutionary perspective?

An implicit evolutionary perspective, or at least a comparative perspective is fundamental to most linguistic research that focuses on the cognitive mechanisms that allow humans to use language. Linguists are not so much interested in *all* cognitive (and anatomical and physiological) machinery that is necessary for language, but are mostly interested in the machinery that appears to be *unique* to humans. This is true for both linguists that propose that language depends on strongly language-specific mechanisms (which are absent in other species because they lack language), and for linguists that propose that language makes use of general cognitive mechanisms (which are supposedly more strongly developed in humans than in other animals for reasons unrelated to language).

The explicit evolutionary perspective has been part of linguistics since Darwin formulated his theory on the origin of species (Darwin, 1859; Darwin & Wallace, 1858) and of the descent of man. An evolutionary perspective on language was

already formulated by Lyell (1863) and Schleicher (1869). An historical review of work on phonology appears in (de Boer, to appear). However for a period, approximately from the 1940s until the end of the 20th century, language was studied as an isolated phenomenon without links to evolution or related behaviors in biology¹. This led to a somewhat contradictory perspective that focused on what made language unique with respect to other species, but that at the same time ignored or downplayed the link with related behaviors in these other species. However, as with any behavior, biology should not be ignored in the study of language. With the words of Theodosius Dobzhansky (1973): "Nothing in biology makes sense, except in the light of evolution." This does not mean that one should always take evolution into account when studying language, but it does mean that if one wants to understand the ultimate causes of language, one should study them from an evolutionary perspective. It also means, as will be explained further in this article, that the evolutionary perspective can help constrain the investigation of (cognitive) adaptations for language. It turns out that evolutionary theory shows that some types of adaptation can evolve, while others are extremely unlikely to have evolved.

The evolutionary perspective can also have practical implications outside the academic study of language and speech: a better understanding of their evolution and their relation to pre-existing behaviors may lead to a better understanding of disorders related to language and speech, and especially of the relation between genetics and disorders. However, the understanding of evolution of language and speech and of their genetics is still too much in its infancy for the potential practical implications to have become usable at the moment.

¹ This is perhaps less true for speech, with for instance, the work of Negus (1938, 1949) and Lieberman (Lieberman & Crelin, 1971; Lieberman, Klatt, & Wilson, 1969) taking an evolutionary perspective. Also, a symposium on the evolution of language was held in 1975 (Harnad, Steklis, & Lancaster, 1976). However, the discussion of evolution of language became more central with the work of Pinker and Bloom (1990) and the evolution of language series of conferences, biannually held since 1996.

The tools of language evolution research

Reconstructing the precise evolution of language is probably impossible; evolution is a historical process and a lot of information of how it happened exactly is lost. However, there is still a lot that can be learned about language evolution, such as: what did our last common ancestor with apes look like, which aspects of modern human anatomy, physiology and cognition have undergone selection related to speech and language, what time course did evolution take and how did evolution happen? The tools that researchers of language evolution bring to bear are partly those of (psycho-) linguistics, but also those of biology (experimental and theoretical), paleontology and cognitive science. All of these provide different windows on language evolution.

In order to study the question of what we share with our (evolutionary) ancestors and under what pressures adaptations for complex communication evolve, comparative biological research is needed. Comparing humans with closely related species (such as chimpanzees, bonobos, gorillas and orangutans) can help establish what we have in common with them. Traits that are shared were most likely already present in the common ancestors with these species (traits that are derived from a common ancestor are called evolutionary *homologues*). Thus it can be established what we inherited from the last common ancestor and what evolved separately after humans separated from the other apes. Fossil evidence acquired by paleontologists can be used to fill in the gaps in the period since the last common ancestor, as well as to get a clearer picture of the ancestral species themselves (which are not necessarily identical to the greatest common denominator of all modern apes).

In comparing apes with humans in this way, it should be noted that the question of whether traits evolved after separation is somewhat orthogonal to the question of whether these traits underwent selection related to language or speech: pre-existing traits can have been fine-tuned for language, while newly evolved traits (that are used in language and speech) may have evolved for different reasons (traits evolved for a different reason than what they are currently used for are called *exaptations* in biology).

Comparative research can also look at groups of less closely related species that show language-like behavior (such as songbirds, bats, pinnipeds and cetaceans,

see Nowicki & Searcy, 2014 for a recent review). Although these behaviors are usually not due to inheritance from a common ancestor (similar traits not evolved from a common ancestor are called evolutionary *analogues*) they may inform us about the circumstances under which these traits evolved, especially if in the group under study the trait is present in some species and not in others. It then becomes possible to figure out which ecological, social, cognitive or other factors predict the evolution of the trait. Of course, no other animal has language, so this approach only makes sense for the study of language evolution if one considers language as a trait that depends on multiple simpler traits (e.g. the ability to produce diverse sounds, the ability to imitate, the ability to associate meanings with signals etc.), some of which can be found in other species. Because of limitations on how evolution can work (explained in more detail below in the section on variation, selection and exaptation) the biological faculty for language must indeed be based on a combination numerous of such simpler traits, rather than on a single monolithic trait.

In addition to comparing observable traits of organisms (biologists call these the *phenotype*) the *DNA* of organisms can also be compared (biologists call this the *genotype*). Because DNA contains a lot of information, it can give a more detailed picture of relations between organisms. Furthermore, it can give estimates of how long ago they started diverging and how diverse the ancient populations must have been. In addition, there are ways to detect whether a gene has undergone positive selection in the recent past. Finally, DNA-researchers have succeeded in isolating and analyzing DNA from fossils (of ancient humans, Neanderthals and equally ancient Denisovans), which provides a more detailed insight into how modern humans are related to their ancestors and how they are different (see Dediu, 2015 for an excellent introduction). What is not possible at the moment is to predict how changes in DNA affect the organisms (i.e. how the genotype affects the phenotype) so it is not possible to reconstruct from DNA how extinct species for which we have recovered DNA were different from modern humans. For instance, DNA evidence cannot (yet) answer the question whether Neanderthal vocal tracts were like modern human vocal tracts.

Instead of looking at biological diversity, it is also useful to look at cultural diversity if one wants to understand the evolution of language and speech. As

will be discussed in the section on limits on evolution below, language evolves through a complex co-evolution between biology and culture. In order to get an idea of cultural evolution and cultural diversity, historical and descriptive linguistics (both of signed and spoken languages) as well as archeological work are important. Relevant data can be derived from such things as the range of variation of modern languages, the speed with which languages change historically and the emergence of new languages (pidgin/creoles or sign languages). However, as this paper is focused on motor control, which is a biologically evolved trait, it will not review this data in detail, as it is mostly related to cultural processes.

The human ability for language evolved (biologically) in the past, and modern human languages evolve (culturally) in ways that cannot be manipulated experimentally. Nevertheless, issues related to language evolution can be studied experimentally. In order to investigate animal abilities (and thus to investigate evolutionary homologies and analogies) experiments with (trained) animals can be done (see Fitch, 2010, chapter 4 for a review). Experiments manipulating DNA have also been performed in order to investigate the role of different genes in vocal learning (examples for FOXP2 are reviewed in Fisher & Scharff, 2010). Cultural evolution of human communication systems is being investigated in controlled experiments (Galantucci, 2009; Scott-Phillips & Kirby, 2010) and is revealing information about human learning biases (Ferdinand, Kirby, & Smith, 2014; van der Ham & de Boer, 2015) and about the dynamics of cultural evolution itself (K. Smith, 2011).

A final tool set to study the evolution of language consists of mathematical and computational models. These may be based on mathematical models from theoretical biology (Blythe & McKane, 2007), on mathematical models from statistical mechanics (Loreto, Baronchelli, Mukherjee, Puglisi, & Tria, 2011) on machine-learning and agent-based models from artificial intelligence (A. D. M. Smith, 2014). These models and the reasons for using them are an important focus of this article, so they will be discussed in some more detail in the next section.

Why computer models?

Computer models are used to help deal with the complexity of evolving systems and with the complexity of the mechanisms (cognitive and anatomical) for dealing with language and speech. This complexity is due to the following factors: the sheer amount of data involved in language, the non-linearity of the dynamic processes involved in evolution and the stochastic nature of evolving systems.

Anything to do with language involves potentially large amounts of data. When dealing with language itself, there are for instance the large amounts of lexical and grammatical information, or the many hours of interaction involved in acquiring a language. When dealing with the history and change of languages, there are for instance the many related languages, dialects and idiolects, the intricate social networks and the rich contextually determined variation of language. When dealing with the mechanisms involved in using language there is the enormous amount of physical, anatomical and physiological detail involved in producing or perceiving it, and of course the staggering complexity of the neural systems involved in processing it. Computers can obviously help to keep track of such data, but *computer models* go beyond this by making it possible to implement a theory as a *computer program*, and then checking whether it actually behaves in the same way as the equivalent real system. Such predictive modeling of language use and acquisition is the subject of computational linguistics, so in this paper we will not go further into systems that only model language acquisition or language use. Models of language evolution go a step further by not only modeling language acquisition and use, but by also modeling social interactions. This makes it possible to "replay" history and thus put models of language change and cultural evolution to the test (e. g. Baxter, Blythe, Croft, & McKane, 2009; Gong, Shuai, & Zhang, 2014).

Similarly, precise models can be made of human vocal anatomy in order to investigate vocal production abilities. There is a long tradition of using computer models of the human vocal tract in order to investigate vocal abilities of other apes (Boë et al., 2017; Fitch, de Boer, Mathur, & Ghazanfar, 2016; Lieberman, Klatt, & Wilson, 1969) where the models are based on observations of existing anatomy, and of our ancestors, where the models are based on reconstructions

on the basis of fossil data (Boë, Heim, Honda, & Maeda, 2002; Lieberman, 2007; Lieberman & Crelin, 1971) or on exploring the effect of hypothetical variations that may correspond to possible ancestral anatomies (Badin, Boë, Sawallis, & Schwartz, 2014; de Boer, 2010).

Computer models do not need to be based on large amounts of data or reproduce every detail accurately in order to be interesting. Many of the processes involved in cultural and biological evolution have complex non-linear dynamics. This has led to a number of authors describing language as a complex dynamical system (Beckner et al., 2009; Kenny Smith, Brighton, & Kirby, 2003; Steels, 2000). In non-technical terms this means that variations on models of language (evolution) such as changes in the social network structure, or in the details of acquisition and (re-) production of language lead to effects that are non-additive, i.e. the effect of variation of two different parameters of the model is not the sum of the effect of each variation individually. Complex in this context means that the long-term behavior of such a system is unpredictable, even if there is no randomness. Small variations in initial conditions may lead to large differences in behavior after only a short period of time. This kind of behavior is called *chaotic* and is found in many natural systems (Fitch, Neubauer, & Herzel, 2002; Hastings & Powell, 1991; Prigogine & Stengers, 1988).

Complex non-linear dynamical systems are notoriously difficult to study with traditional mathematical techniques, but they are quite straightforward to study with computer models. Hence computer models have been extensively used in the study of the evolution of language and speech, especially in exploring the dynamics of cultural evolution and its interaction with biological evolution. Examples will be discussed in the section on co-evolution below.

The final source of complexity is the randomness of evolutionary processes: while survival (both of biological and cultural entities) is partly determined by their quality, coincidence also plays an important role. Randomness can of course be modeled quite straightforwardly in computer simulations through the use of functions that can generate (pseudo-) random numbers. However, the role of coincidence in evolutionary processes has also been studied for a long time in theoretical biology and in statistical physics (Blythe & McKane, 2007; Crow & Kimura, 1970; Loreto et al., 2011; van Kampen, 1992). These are mostly

mathematical models that make concrete predictions about average properties that would be observed if an evolutionary process occurs many times. The resulting equations can be solved analytically for some simple situations, but for more complex (more realistic) situations, an analytic solution may not be obtainable. However, numerical software can effectively calculate results for a large class of these mathematical models; this approach has for instance been used in calculating whether co-evolution between language and biological adaptations for language is likely [ref?].

Mathematical models tend to be limited to simpler, less realistic cases than direct computer simulations. In order to do mathematical analysis, the system under study must be converted to a particular set of equations that fits the available analysis techniques. For a direct simulation, one is much freer to formulate the behavior of the system, the only requirement being that it can be translated into a computer program. However, mathematical analysis (even if it is through numerical approximation) tends to be faster and provide more insight in the influence of parameters and initial conditions of a system than computer simulation. Thus mathematical and computational modeling are complementary in the contributions they can make.

Differences between Apes and Humans

In order to understand what aspects of the human ability for speech and language may have undergone selection that was actually related to speech or language, it is useful to compare humans with their closest evolutionary relatives, the great apes. By figuring out what traits (related to speech) this group has in common, and which traits in humans are different compared to the other members of the group, it is possible to reconstruct the last common ancestor of the great apes, and thus to have an idea which traits have been modified (possibly because of selection related to speech) in humans. In doing this, it should be kept in mind that our last common ancestor was not exactly like any of the extant great apes; all great apes have been shaped by the intervening millions of years of evolution. However, traits that are shared by all great apes,

but not by humans, are nevertheless likely to have been present in the last common ancestor.

Behavioral differences

In the wild, the most obvious difference in spontaneous behavior is of course that all humans have language (spoken for the hearing, and signed for the deaf) while other great apes do not. However, there are other, subtler differences and similarities that may be of relevance. Apes (i.e. orangutans, gorillas, bonobos and chimpanzees) do appear to communicate with gestures (Pika, Liebal, Call, & Tomasello, 2005) although these gestures appear to be much more limited than human ones. They tend to be dyadic rather than triadic gestures (e.g. gestures to draw the attention of another ape, rather than gestures about a third party) and they tend to be idiosyncratic and individual rather than conventionalized in the group. Although apes in the wild only appear to communicate vocally with limited repertoires of innate calls, chimpanzees (Crockford, Herbinger, Vigilant, & Boesch, 2004; Watson et al., 2015) and orangutans (Hardus, Lameira, van Schaik, & Wich, 2009) do appear to have vocal culture, in the sense that each group's calls differ in their precise characteristics from equivalent calls in a different group. This fits in a broader observation that (at least) chimpanzees (Whiten, 2017; Whiten et al., 1999) and orangutans (Krützen, Willems, & van Schaik, 2011; van Schaik et al., 2003) have culture to the extent that certain learned behaviors (such as nut cracking) are transmitted from generation to generation. As far as motor control and motor learning are concerned, it appears that apes have some ability in this area (Whiten, 2017), but that their learning is more guided by the goal of the action than by precise copying of the actions. For instance, in learning how to crack nuts, chimpanzees appear to adapt their nut cracking more to individual situations, while human hunter-gatherers use a perfected (but less flexible) cooperative system (Boesch, Bombjaková, Boyette, & Meier, 2017). Although this is of course not quite the same as fine control over movements of vocal articulators, it has been proposed that vocal motor control in humans is closely related to motor control of manual movements (Calvin, 1993). Finally, apes appear to be less altruistic, and less (if at all) inclined to

share information with others, whereas humans (especially infants) appear to have an innate desire to share information (Fitch, 2010, section 3.8.8).

Extrapolating this to the last common ancestor of other great apes and humans, it appears that it may have had rudimentary culture and cultural learning, and some degree of gesturing, vocal adaptation and motor learning. However, there is little evidence that the last common ancestors shared information voluntarily or communicated vocally for communication's sake. All of these skills have been greatly enhanced in modern humans, and are candidates for behaviors that have undergone positive selection related to speech and language.

Anatomical differences

Beside the behavioral differences between humans and other great apes, there are also a number of anatomical differences that could be of relevance to the study of speech. These can be used to reconstruct traits of the last common ancestor in the same way as for behavioral traits. Moreover, even though the anatomical structures involved in speech mostly consist of soft tissue, which does not fossilize well, a number of (bony) fossil indicators can be used for this reconstruction as well.

The best-studied difference between other great apes and humans is the position of the larynx (and the related differences in tongue shape). There are other differences such as the presence or absence of air sacs, control over breathing, control of the tongue, control of the vocal folds and the shape of the chin. All of these are potential candidates for having undergone selection related to speech.

The position of the larynx has been investigated in the light of evolution since at least Negus (1938), and was the topic of what probably was the first application of computer models to questions about the evolution of speech (Lieberman et al., 1969). The human larynx is positioned lower with respect to the oral cavity in humans than it is in other great apes. The hypothesis is that a lower larynx (and the rounder tongue shape that this entails) leads to more articulatory flexibility, and thus to a larger range of possible speech sounds, because it allows for independently controlling the size of the oral and pharyngeal cavities. Although the question of whether a lowered larynx has evolved for speech has been hotly debated (Boë et al., 2002; Fitch, de Boer, Mathur, & Ghazanfar, 2017; Lieberman,

2007, 2017) – and conflated with the independent questions of whether Neanderthals had lowered larynges and whether a lowered larynx is an essential precondition for speech – modern consensus is that although the modern human larynx position leads to somewhat greater vocal flexibility, even standard primate anatomy allows for a sufficiently large range of speech sounds for language (Boë et al., 2017; Fitch et al., 2016).

Air sacs have also been investigated as a potential indicator of the evolution of speech. Air sacs (at least the ones that occur in great apes) are large pouches (some can hold multiple liters of air, although in all species that have them there is considerable individual variation) that are connected to the vocal tract through openings that occur just above the vocal folds. Although the idea that the loss of air sacs was important for language evolution has been proposed already at least since (Fitch, 2000) their acoustic effects have been investigated more recently (de Boer, 2008, 2009; Riede, Tokuda, Munger, & Thomson, 2008) using a combination of acoustic analysis, computer reconstructions and an experimental perceptual study, leading to the conclusion that their presence influences production of clear speech (de Boer, 2012). Moreover, this has been linked to fossil evidence, leading to the hypothesis that air sacs were still present in Australopithecines about 3 million years ago, and had disappeared in *Homo heidelbergensis* and the last common ancestor of *Homo sapiens* and Neanderthals at approximately 400 000 years ago. From this it follows that complex vocalizations may have started evolving somewhere between these dates.

Another important difference between humans and other great apes turns out to be control over breathing. In speech humans produce very long and controlled outbreaths, and very rapid inbreaths, with vocalization occurring almost exclusively on the outbreath. MacLarnon and Hewitt (2004) have proposed that this control over breathing correlates with the size of the *thoracic vertebral canal*, through which the nerves pass that control the intercostal muscles – the ones involved in controlling breathing in speech. They found (on the basis of measurements of vertebrae from both recent and fossil origins) that the canal is enlarged in humans and Neanderthals, but not in chimpanzees or *Homo ergaster*. From this they conclude that breathing control and hence complex vocalizations

must have evolved before the last common ancestor with Neanderthals. This is in line with the evidence gleaned from air sacs.

There are more anatomical differences between modern humans, ancestral hominins and other great apes, but these turn out to be unrelated to speech, have been less well investigated, or there is as yet no satisfactory explanation for the differences. One example where there is no apparent relation with speech is the canal through which the hypoglossal nerve passes. This is one of the nerves that controls tongue movement. Analogous to the thoracic vertebral canal, it has been proposed (Kay, Cartmill, & Balow, 1998) that the size of the hypoglossal canal may tell us something about accuracy of tongue motor control. However, it turns out that canal size does not reliably predict nerve size, and that hypoglossal canals in humans are not significantly different from those of other apes (DeGusta, Gilbert, & Turner, 1999). Moreover, imaging studies in living monkeys (*Macaca fascicularis*) and in humans have shown that tongue movements involved in eating are as complex (or more complex) than those involved in talking (Hiimae, Hayenga, & Reese, 1995; Hiimae & Palmer, 2003) making it unlikely that more innervation of the tongue needed to evolve for speech.

One potential difference between humans and other great apes that has not been well-investigated is that of the vocal folds. These are quite different in chimpanzees for instance, but very little is known about what these differences in anatomy mean for their function. Finally, the chin is an example where there are clear differences: modern humans are the only ape with a chin (i. e. a bony protrusion). Even Neanderthals did not have this. Although there has been speculation for a long time that this may be related to speech (Schwartz & Tattersall, 2000), possibly because it allows for more freedom of motion of the tongue, there is as yet no satisfactory explanation and therefore no consensus about the function of the human chin or its role in speech.

Cognitive differences

Not all differences between humans and other great apes are easily observed in the wild or through examination of their anatomy. Some differences (and similarities) have only become clear through experiments in a controlled setting.

This is especially true for (cognitive) abilities related to fine control over breathing and vocalization.

Early research with trained apes was somewhat haphazard and not very well controlled, but it did show that chimpanzees raised in a human family were unable to learn how to speak, although in other aspects their development was similar to that of a human child. The earliest such experiment was with Gua (Kellogg & Kellogg, 1933), and this was repeated for Vicki (Hayes & Hayes, 1951). One reason why the chimpanzee babies did not learn to speak was that they were not able to control their vocal folds voluntarily (while Vicki did learn to control her tongue and lips in a rudimentary way for some words, she did not learn to control her vocal folds). Later research focusing on sign language showed that apes are in fact able to learn and use sizeable lexicons, i.e. signal-meaning mappings (Gardner & Gardner, 1969; Patterson, 1978; Savage-Rumbaugh, McDonald, Sevcik, Hopkins, & Rubert, 1986; Savage-Rumbaugh et al., 1993). The cognitive limitations preventing acquisition of speech of our evolutionary ancestors must therefore have been related to control over the vocal folds (and possibly the vocal tract and breathing) rather than an inherent inability to relate signals to meanings.

There are of course many other cognitive differences between humans and other great apes that are relevant for language. Especially differences related to social faculties, such as theory of mind, the ability and willingness to cooperate, the ability for ostensive inferential communication and of course the ability to use phonological or syntactic structure have been the focus of research (for a review in the context of the evolution of speech, see Fitch, 2010, chapter 4). All these abilities are present in humans to a much greater extent than in other great apes, and therefore most likely have undergone evolution in the time since the last common ancestor. Nevertheless, most of these abilities are present to some degree in other great apes, while it is also true that the details of the experimental design strongly influence the outcome (Leavens, Bard, & Hopkins, 2017).

Similarly, it appears that control over the vocal folds and an ability to imitate and voluntarily control some utterances is not entirely lacking in other great apes. Vicki did learn to say a few words (without involvement of the vocal folds).

There is evidence of vocal culture and vocal adaptation in chimpanzees, both in the wild (Crockford et al., 2004) and in captivity (Watson et al., 2015, but their finding is controversial). More anecdotally, orangutans in captivity have been observed to produce whistles and raspberries (Lameira, Hardus, Mielke, Wich, & Shumaker, 2016; Wich et al., 2009) while the trained gorilla Koko has been observed to produce vocal and play behaviors that appear to involve glottal stops and glottal noise (de Boer & Perlman, 2014), implying at least rudimentary voluntary control over the vocal folds. Hence vocal control and motor control over the articulators did not appear entirely *de novo* in human evolution since the last common ancestor with the other great apes.

Implications for Speech Motor Control

Comparative research thus reveals that although modern human anatomy has been modified to some extent to allow efficient production of a larger range of speech sounds, the major innovations must have been cognitive. The last common ancestor of humans and great apes must have had a vocal tract that was in principle able to produce a range of sounds that is sufficient for understandable speech. However, it lacked sufficiently precise voluntary control, especially of the vocal folds, to do this. In addition, compared to modern humans, it must have had much reduced abilities for learning, imitation and for social reasoning to allow for a learned, open communication system like human language. Moreover, it must have lacked the ability to learn to analyze speech (and language) in terms of a limited set of building blocks that can be recombined into an unlimited number of utterances. All these abilities must have evolved in the roughly seven million years that separate modern humans from this last common ancestor.

Limits on Evolution

Comparative research can determine the differences between the last common ancestor and modern humans, and thus provide a list of traits that have evolved in modern humans. Still, not all of the traits that are different and that are now used in language and speech need to have evolved under selective pressure for this purpose. It is always possible that they evolved for other reasons, or are

accidental byproducts of unrelated evolutionary changes. In order to understand what may have evolved for language and speech, it is essential to understand the theory of the evolutionary process.

Variation, Selection and Exaptation

An important point to understand about evolution is that it needs to work with what is already there. Evolution works by selecting variations, some of which may result in higher fitness (the ability to produce offspring). This is called positive selection. If a trait is bad, it results in negative selection (fewer offspring). Traits under positive selection tend to increase in frequency over time, while traits under negative selection tend to disappear. However, in small populations, random effects (called drift) are also important, meaning that sometimes traits can evolve that give no fitness advantage, or that may even be (slightly) detrimental.

In the large majority of cases, the magnitude of variations tends to be small. Variations are eventually caused by random changes in the DNA (mutations). The probability of such variations to have a negative effect is much larger than to have a positive effect, and the probability of a large positive effect is vanishingly small (Eyre-Walker & Keightley, 2007; Orr, 2003). Therefore, positive selection works on traits that are small variations of what was already present in previous generations. Large changes occur through accumulation of smaller changes over time. There are two important consequences of this.

The first consequence is that any complex trait, for instance the ability to learn structured speech, the ability to imitate or the anatomy of the vocal tract, is the result of the accumulation of many small changes in the DNA. This means that for any such trait, many genes are involved, and hence that the relation between the genes (the genotype) and anatomy or behavior (the phenotype) is far from straightforward.

The second consequence is that any trait is a modification of a pre-existing one, and this pre-existing trait has at some point back in time evolved for other purposes. An example would be the larynx: this originally evolved to be able to close off the airway (Negus, 1949) but is now essential for producing speech.

The fact that any trait related to speech and language builds on pre-existing traits also means that the question of whether such a trait is "language-specific" or "domain-general" is meaningless in evolutionary terms². The answer to this question depends on one's definition of what these terms mean, so there is no objective answer to them. The only question that can be answered (but almost never in a simple way) is whether a trait has undergone selection related to speech or language.

Traits involved in a given behavior (such as language or speech) may have undergone different amounts of selection related to that behavior. On one extreme, a trait may have undergone a lot of selection for a long time, and then one would expect that trait to be specialized in certain ways for that behavior. For instance, human legs and feet are strongly optimized for bipedal walking, making it very difficult for humans to walk on all fours. On another extreme, a trait involved in a behavior may have undergone no selection at all related to that behavior. In general such a trait is called an *exaptation*. Some accounts of language evolution propose that all traits involved in language are exaptations (e. g. Bolhuis, Tattersall, Chomsky, & Berwick, 2014), but also (Tomasello, 1995). The facts that variations tend to be small and that traits are based on pre-existing traits that evolved for a different purpose, make it very difficult to detect whether any trait has undergone selective pressure related to a given function, let alone for a function as complex, ephemeral and abstract as language or speech. However, there are a number of indications. Optimality for a given function is one such indication (Parker & Maynard Smith, 1990). If it turns out that changing a trait (such as larynx height) in any way reduces a function related to that trait (such as the range of speech sounds that can be produced) that is a strong indication that that trait has undergone selection related to that function. This has been used to argue that the modern human larynx position has undergone selection related to speech (de Boer, 2010), but see (Badin et al., 2014) for a critical analysis. Another indication is specialization: if a certain trait,

² These terms may however be meaningful in describing function: a mechanism may be domain-specific in the sense that it works better in a certain domain than in others, or domain-general in that it works equally well in multiple domains.

which could in principle be general, shows specialization for a given domain, this is an indication that it has undergone selection for that domain. For instance Cosmides and Tooby (1992) have shown that certain cognitive tasks are much easier for human participants when presented as a social problem (i.e. cheater detection) than as an abstract puzzle. They argue that this indicates that these cognitive abilities therefore evolved under selection for these social functions, which are indeed important in primate groups.

In the context of speech it could be investigated whether certain tasks are easier with speech sounds than with other signals (van der Ham & de Boer, 2015), or whether humans are more sensitive to speech than to other acoustic signals, or have speech-specific brain areas (e. g. Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). In the context of language, there is a lively debate about whether the cognitive machinery for language is specialized or domain-general in this way (e.g. Evans & Levinson, 2009 and the accompanying commentaries).

Finally, DNA research can detect signals of (positive or negative) selection in the genome (see Dediu, 2015 section 8.5.6 for an introduction). Although this does not link indications of selection to traits directly (because the link between genes and traits is in general not direct) it does allow checking whether genes that have in some way been linked to language or speech have undergone recent positive selection. This has for instance been done for FOXP2 (Enard et al., 2002).

These restrictions on how traits vary and change make it difficult to study the evolution of those traits. However, they do help to constrain how evolution can take place, because each new variation needs to spread through the population, and this is only likely to happen if the trait increases the fitness of an organism.

Fitness and Spread of Variations

The fitness of an organism is by definition equal to the number of offspring it produces. The fitness of a *trait* is the average number of offspring that individuals *with that trait* produce. If this is higher than the average of the whole population, then that trait is likely to spread through the population. However, there can be large variation in the number of offspring for individuals with identical traits; even potentially fit individuals may end up without offspring by accident. This means that if a trait is very rare, it is quite possible that it dies out,

even though it would in principle result in higher-than-average fitness. Because new variations emerge through random mutations that occur in only one individual at a time, they are always rare initially, and therefore random effects need to be taken into account in addition to fitness effects. Random effects are especially important in small populations, because the effect of random fluctuations tends to average out in larger populations.

Random fluctuations (called *drift* by biologists) have an effect that is relevant for the study of traits related to language (and indeed of any trait). They cause some variants to disappear from the population, even though they may not be less fit than other variants. Indeed, because differences between variants tend to be small, and therefore generally do not cause large differences in fitness, it is likely that many potentially beneficial variants never make it. Consequently, it is likely that the traits that survive eventually are not the best possible traits, and that some traits are the way they are through coincidence. This means that one should not necessarily assume that all observed traits can be explained through adaptation (i.e. they spread because they conveyed higher-than-average fitness). On the other hand, it is extremely unlikely that a complex trait (one influenced by many genes) did not undergo selection related to its function at all. Although some of the contributing genes may have come to be in their present state through drift or through selection for other functions, if a complex trait is to be specialized at all for its function, many of the contributing genes must have undergone selection related to this function. Because beneficial mutations are rare, it is unlikely that the many gene variants involved emerged at the same time. Therefore, each of these gene variants must by themselves have made a positive contribution to the trait's fitness. This is in general true for the evolution of any complex trait, including language and speech: each contribution to the trait must have increased fitness. Hence evolution of complex traits happens along a "*path of ever increasing fitness*" (Zuidema, 2005). This is something that needs to be kept in mind when proposing scenarios for the evolution of a complex trait: not only does the eventual (complete) complex trait need to provide a fitness benefit, intermediate (incomplete) stages each also need to provide a fitness benefit.

Finding a path of ever increasing fitness is particularly challenging for the evolution of language and speech. This is because fitness related to language and speech may be frequency-dependent. Frequency-dependent fitness is fitness that does not only depend on the trait itself, but also on the frequency with which it occurs in the population. It can be negative, in which case a higher frequency in the population results in lower fitness, but in the case of language, frequency-dependency is *positive*: because language and speech are tools for communication, the more an adaptation for them is shared by members of a population, the more beneficial this adaptation is.

Consequently, when a variant trait that would in principle be beneficial for language and speech first arises in a single individual, it may not confer any fitness advantage at all. This is a potential problem for any account of the evolution of language: what is the use of a mutation for language if no one else in the population shares it? In order to counter this problem, it has been proposed that adaptations for language first evolved for other purposes (most notably abstract thought, e. g. Bolhuis et al., 2014) or that language does not depend on traits that have evolved for it at all (Tomasello, 1995). In principle it is also possible that traits first spread through drift in small (sub-) populations, and thus reached a critical frequency so that they could undergo positive selection. However, if one takes into account that traits for using language do not evolve (biologically) in isolation, but that language itself evolves culturally, it turns out that frequency-dependent fitness becomes much less of a problem. One therefore needs to take into account co-evolution between biology and culture when studying the evolution of language and speech.

Culture-biology co-evolution

Even though humans are not the only species that have culture, and whose biological evolution is influenced by culturally transmitted³ behavior, they are the species for whom the interaction between biology and culture is most important. This is true for culturally transmitted knowledge about for instance

³ Culturally transmitted behaviors are not just learned behaviors; they are behaviors that are learned by observing *others* performing that behavior.

agriculture, cooking or technology, all of which have resulted in biological adaptations (see Dediu, 2015 chapter 9 for examples) but also for language and speech.

Cultural evolution can cause language and speech to become optimized for its function, just as biological evolution can optimize biological traits for their function. In certain cases cultural evolution can lead to emergence of structure (e. g. de Boer, 2000), or emergence of complexity (Kenny Smith et al., 2003; Zuidema & de Boer, 2009). In this way, cultural evolution can set up conditions for biological evolution to work on, thus alleviating the problem of frequency-dependency in the evolution of biological adaptations for language. Cultural evolution can, given existing biological constraints on for instance articulation or language learning, create a culturally transmitted language of a complexity that is somewhat challenging to learn and use for the members of the population. Biological variants that help to deal with the culturally transmitted language can then spread, without the need for these variants to be shared by others. After all, the rest of the population can already deal with the complexity of the language, albeit with slightly more effort. When the more capable variant has become frequent in the population, cultural evolution may push the language to become even more complex. Thus a co-evolution between biology and culture may cause an increase in the complexity of both the language itself and of the mechanisms to learn, perceive and produce that language. This process has been demonstrated (using a computer simulation) for interaction between the biological evolution of the vocal tract and the cultural evolution of vowel systems (de Boer, 2016).

It turns out that co-evolution between biology and culture does not always work in this way. If the traits under evolution are learning biases for discrete traits (such as word order) it turns out that the interaction between culture and biology actually prevents strong biological preferences (called *biases* in the literature) for one (cultural) variant to evolve (Thompson, Kirby, & Smith, 2016). This is because cultural evolution tends to amplify the effect of biological (cognitive) biases. Therefore, once a small biological bias exists that aids learning a culturally transmitted language, an increase in the strength of that bias does not confer extra fitness. This, in combination with mutational pressure that tends

to reduce the strength of the bias back to a neutral value, prevents strong biases for such discrete culturally transmitted traits to evolve. In this case, co-evolution between culture and biology does not exactly prevent biases for language learning to evolve, but it causes biases that evolve to be weak and *defeasible* (i.e. given enough counterexamples, individuals remain able to learn a language that violates the bias). The practical consequence of this is that evidence for language-related selection may be hard to detect experimentally, because such weak biases will only cause small empirically observable effects.

Another obstacle that has been proposed against the evolution of biological adaptations for learning language is the fact that culture tends to change faster than biology, and that therefore (culturally transmitted) language may not present a sufficiently stable target for biological evolution to adapt to (Chater, Reali, & Christiansen, 2009). Although it is true that rapidly changing culture may reduce the probability of biological adaptations evolving, it turns out that the original agent-based model on which the conclusions about the problem of rapid cultural change were based was heavily biased against such adaptations. A mathematical model has shown that even if culture changes rapidly, biological adaptations *can* evolve, especially when having a biological preference for a certain type of language increases the probability of that language arising culturally (de Boer & Thompson, 2018).

When studying the evolution of biological adaptations for language and speech, it is therefore essential to take into account that language(s) also evolve culturally. Depending on the details of the trait that evolves and on the nature of the interaction between the biological properties of the individuals and the culture that arises in a population of these individuals, cultural evolution may enhance or reduce the effects of biological evolution. Moreover, the fact that biological evolution occurs in the context of a culturally evolving language restricts the kinds of adaptations that can evolve. Understanding co-evolution between biology and culture is therefore essential for understanding evolution of language and speech. Computational and mathematical models have played an essential role in increasing this understanding.

Implications for Speech Motor Control

The fact that the evolution for adaptations for speech motor control (if there are such adaptations) must fulfill certain constraints has consequences for how speech motor control manifests itself in modern humans. One of the constraints is that evolution must have happened in many small steps, and from this it follows that there cannot be a simple genetic mechanism for speech. Another constraint is that modern adaptations for speech must be modifications of pre-existing traits. As we can reconstruct some of those pre-existing traits by looking at what we have in common (or not) with other great apes, it becomes clear that although there has probably been some fine-tuning of the vocal tract, the major changes must have been of our ability to voluntarily control the vocal folds and the vocal tract and to learn to sequence articulatory movements. Whether these abilities underwent selection related to speech and language remains to be demonstrated, although it seems unlikely that they evolved through pure random drift.

The fact that biological adaptations must have co-evolved with culturally evolving languages puts other constraints of what could have evolved. On the one hand, this may have helped to solve the problem of frequency-dependence: every small step must have conveyed a fitness advantage, but in a naive view such a fitness advantage would not have existed in the case where the adaptation was rare, and thus not shared with others in the population. However, it turns out that in certain cases cultural evolution causes language to be challenging for members of the population, such that biological adaptations convey an advantage, even if only one individual in the population has them.

The cases in which this process works, are the ones where both the property of the language and the biological trait in question are open-ended, for instance the extent of a vowel system and the position of the larynx, the complexity of words and the capacity of the phonological loop memory or the number of phonological distinctions in a language and the accuracy of motor control. The cases in which this process does not work are the ones where linguistic traits are discrete, such as whether a language has final devoicing or not. In this case, small biases are likely to evolve, but these will be defeasible, and hard to detect experimentally.

This may help to identify promising areas for empirical research into adaptations for language and speech.

Conclusion

Evolution of speech and language is hard to study. The first reason is that language and speech, and their evolution are inherently complex topics to study. The second reason is that the study of evolution is a historical science, and a lot of information has been lost because neither language nor speech themselves, nor the soft anatomical structures that help generate them leave fossil traces. However, it turns out that with the techniques reviewed above quite a lot can be learned about what evolved, how this has happened and partly when this has happened.

Comparison with other species, the use of fossils and reconstructions based on computer models allows for establishing properties of our last common ancestor, and from there allow for an estimate of what may have evolved, and in some cases approximately when. The emerging consensus view is that adaptations for speech and language (or at least for complex vocal behavior such as song) predate the last common ancestor with Neanderthals who lived approximately 400 000 years ago. The adaptations consisted of fine-tuning of the anatomy of the vocal tract, but it appears that the major, and probably the earlier adaptations must have had to do with vocal (motor) control: control over breathing, control over the vocal folds and the ability to imitate complex vocalizations and to analyze these in terms of a small set of building blocks.

Theoretical and computer modeling studies of the evolutionary process that take into account the fact that (culturally transmitted) language(s) and biological adaptations for language co-evolve, allow us to predict more or less what kind of adaptations to expect. For discrete traits (such as a preference for final devoicing) we do not expect to find strong adaptations, but for open-ended continuous traits, such as the position of the larynx or the accuracy of motor control, we expect stronger adaptations. The former may be hard to detect empirically, the latter may be easier to detect. However, there appears to be no hard evidence for selection related to language and speech, yet, except possibly

for the position of the larynx. Such evidence is expected to take the form of either optimality for a language- or speech-related function, or specialization for speech and language in the sense that the trait under study performs better for speech- and language related tasks than for similar tasks in a different domain.

The study of evolution of speech and language in its modern form is hardly older than 20 years, while the insight that co-evolution between biology and culture is critical is perhaps less than ten years old. The field is therefore still in its infancy, and much still needs to be learned, both about the relevant facts and about the methods that can be used to investigate them. Models, both computational and mathematical, are important in this context in order to deal with the complexity of the data and the complexity of the dynamics of the evolutionary process.

References

- Badin, P., Boë, L.-J., Sawallis, T. R., & Schwartz, J.-L. (2014). Keep the lips to free the larynx: Comments on de Boer's articulatory model (2010). *Journal of Phonetics*, 46, 161–167.
- Baxter, G. J., Blythe, R. A., Croft, W., & McKane, A. J. (2009). Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, 21(02), 257–296.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, P. T. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, 59(Suppl. 1), 1–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309.
- Blythe, R. A., & McKane, A. J. (2007). Stochastic models of evolution in genetics, ecology and linguistics. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07), P07018.

- Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., ... Fagot, J. (2017). Evidence of a Vocalic Proto-System in the Baboon (*Papio papio*) Suggests Pre-Hominin Speech Precursors. *PloS One*, *12*(1), e0169321.
- Boë, L.-J., Heim, J.-L., Honda, K., & Maeda, S. (2002). The potential Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics*, *30*(3), 465–484.
- Boesch, C., Bombjaková, D., Boyette, A., & Meier, A. (2017). Technical intelligence and culture: Nut cracking in humans and chimpanzees. *American Journal of Physical Anthropology*, *163*(2), 339–355.
- Bolhuis, J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How Could Language Have Evolved? *PLOS Biology*, *12*(8), e1001934.
- Calvin, W. H. (1993). The unitary hypothesis: A common neural circuitry for novel manipulation, language, plan-ahead and throwing? In K. R. Gibson & T. Ingold (Eds.), *Tools, language and cognition* (pp. 230–250). Cambridge: Cambridge University Press.
- Chater, N., Reali, F., & Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences*, *106*(4), 1015–1020.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). Oxford: Oxford University Press.
- Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild Chimpanzees Produce Group-Specific Calls: a Case for Vocal Learning? *Ethology*, *110*, 221–243.

- Crow, J. F., & Kimura, M. (1970). *An Introduction to Population Genetics Theory*.
Caldwell, NJ: The Blackburn Press.
- Darwin, C. (1859). *On the origin of species*. reprinted Penguin Classics, 1985.
- Darwin, C., & Wallace, A. R. (1858). On the tendency of species to form varieties;
and on the perpetuation of varieties and species by natural means of
selection. [Read 1 July]. *Journal of the Proceedings of the Linnean Society of
London. Zoology*, 3, 45–50.
- de Boer, B. (to appear). Phonology and Evolution. In B. E. Dresher & H. G. van der
Hulst (Eds.), *The Oxford Handbook of the History of Phonology*. Oxford:
Oxford University Press.
- de Boer, B. (2000). Self organization in vowel systems. *Journal of Phonetics*,
28(4), 441–465.
- de Boer, B. (2008). The acoustic role of supralaryngeal air sacs. *Journal of the
Acoustical Society of America*, 123(5 Pt 2), 3732–3733.
- de Boer, B. (2009). Acoustic analysis of primate air sacs and their effect on
vocalization. *Journal of the Acoustical Society of America*, 126(6), 3329–
3343.
- de Boer, B. (2010). Investigating the acoustic effect of the descended larynx with
articulatory models. *Journal of Phonetics*, 38(4), 679–686.
- de Boer, B. (2012). Loss of air sacs improved hominin speech abilities. *Journal of
Human Evolution*, 62(1), 1–6.
- de Boer, B. (2016). Modeling co-evolution of speech and biology. *Topics in
Cognitive Science*, 8(2), 459–468.

- de Boer, B., & Perlman, M. (2014). Physical mechanisms may be as important as brain mechanisms in evolution of speech. *Behavioral and Brain Sciences*, 37(5).
- de Boer, B., & Thompson, B. (2018). Biology-culture co-evolution in finite populations. *Scientific Reports*, 8(1), 1209.
- Dediu, D. (2015). *An Introduction to Genetics for Language Scientists: Current Concepts, Methods and Findings*. Cambridge: Cambridge University Press.
- DeGusta, D., Gilbert, W. H., & Turner, S. P. (1999). Hypoglossal canal size and hominid speech. *Proceedings of the National Academy of Sciences*, 96(4), 1800–1804.
- Dobzhansky, T. (1973). Nothing in biology makes sense, except in the light of evolution. *The American Biology Teacher*, 35, 125–129.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., ... Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418, 869–872.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32, 429–492.
- Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8), 610.
- Ferdinand, V., Kirby, S., & Smith, K. (2014). Regularization in Language Evolution: On the Joint Contribution of Domain-Specific Biases and Domain-General Frequency Learning. In E. A. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), *The Evolution of Language: Proceedings of the 10th International Conference (EVO LANG10)* (pp. 435–436). Hackensack, NJ: World Scientific.

- Fisher, S. E., & Scharff, C. (2010). FOXP2 as a molecular window into speech and language. *Trends in Genetics*, 25(4), 166–177.
- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, 4(7), 258–267.
- Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.
- Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2(12). <https://doi.org/10.1126/sciadv.1600723>
- Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2017). Response to Lieberman on “Monkey vocal tracts are speech-ready.” *Science Advances*, 3(7). <https://doi.org/10.1126/sciadv.1701859>
- Fitch, W. T., Neubauer, J., & Herzog, H. (2002). Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, 63(3), 407–418.
- Galantucci, B. (2009). Experimental Semiotics: A New Approach for Studying Communication as a Form of Joint Action. *Topics in Cognitive Science*, 1, 393–410.
- Gardner, R. A., & Gardner, B. T. (1969). Teaching Sign Language to a Chimpanzee. *Science*, 165(3894), 664–672.
- Gong, T., Shuai, L., & Zhang, M. (2014). Modelling language evolution: Examples and predictions. *Physics of Life Reviews*, 11(2), 280–302.
- Hardus, M. E., Lameira, A. R., van Schaik, C. P., & Wich, S. A. (2009). A description of the orangutan’s vocal and sound repertoire, with a focus on geographic

- variation. In S. A. Wich, T. Mitra Setia, S. S. Utami, & C. P. van Schaik (Eds.), *Orangutans* (pp. 49–60). New York: Oxford University Press.
- Harnad, S. R., Steklis, H. D., & Lancaster, J. (1976). *Origins and Evolution of Language and Speech*. New York (NY): New York Academy of Sciences.
- Hastings, A., & Powell, T. (1991). Chaos in a Three-Species Food Chain. *Ecology*, 72(3), 896–903.
- Hayes, K. J., & Hayes, C. (1951). The intellectual development of a home-raised chimpanzee. *Proceedings of the American Philosophical Society*, 95(2), 105–109.
- Hiiemae, K. M., Hayenga, S. M., & Reese, A. (1995). Patterns of tongue and jaw movement in a cinefluorographic study of feeding in the macaque. *Archives of Oral Biology*, 40(3), 229–246.
- Hiiemae, K. M., & Palmer, J. B. (2003). Tongue Movements in Feeding and Speech. *Critical Reviews in Oral Biology & Medicine*, 14(6), 413–429.
- Kay, R. F., Cartmill, M., & Balow, M. (1998). The hypoglossal canal and the origin of human vocal behavior. *Proceedings of the National Academy of Sciences*, 95, 5417–5419.
- Kellogg, G. N., & Kellogg, L. A. (1933). *The Ape and The Child: A Comparative Study of the Environmental Influence Upon Early Behavior*. New York, NY: Hafner Publishing Co.
- Krützen, M., Willems, E. P., & van Schaik, C. P. (2011). Culture and Geographic Variation in Orangutan Behavior. *Current Biology*, 21, 1808–1812.
- Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A., & Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific Reports*, 6, 30315.

- Leavens, D. A., Bard, K. A., & Hopkins, W. D. (2017). The mismeasure of ape social cognition. *Animal Cognition*, 1–18.
- Lieberman, P. H. (2007). Current views on Neanderthal speech capabilities: A reply to Boë et al. (2002). *Journal of Phonetics*, 35(4), 552–563.
- Lieberman, P. H. (2017). Comment on “Monkey vocal tracts are speech-ready.” *Science Advances*, 3(7). <https://doi.org/10.1126/sciadv.1700442>
- Lieberman, P. H., & Crelin, E. S. (1971). On the speech of Neanderthal man. *Linguistic Inquiry*, 2, 203–222.
- Lieberman, P. H., Klatt, D. H., & Wilson, W. H. (1969). Vocal Tract Limitations on the Vowel Repertoires of Rhesus Monkey and other Nonhuman Primates. *Science*, 164, 1185–1187.
- Loreto, V., Baronchelli, A., Mukherjee, A., Puglisi, A., & Tria, F. (2011). Statistical physics of language dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(04), P04006.
- Lyell, C. (1863). *The Geological Evidences of the Antiquity of Man*. London: John Murray.
- MacLarnon, A., & Hewitt, G. P. (2004). Increased Breathing Control: Another Factor in the Evolution of Human Language. *Evolutionary Anthropology*, 13, 181–197.
- Negus, V. E. (1938). Evolution of the Speech Organs of Man. *Archives of Otolaryngology*, 28, 313–328.
- Negus, V. E. (1949). *The Comparative Anatomy and Physiology of the Larynx*. London: William Heinemann Medical Books Ltd.
- Nowicki, S., & Searcy, W. A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology*, 28, 48–53.

- Orr, H. A. (2003). The distribution of fitness effects among beneficial mutations. *Genetics*, 163(4), 1519–1526.
- Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, 348, 27–33.
- Patterson, F. G. (1978). The gestures of a gorilla: Language acquisition in another pongid. *Brain and Language*, 5(1), 72–97.
- Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). The gestural communication of apes. *Gesture*, 5(1/2), 41–56.
- Pinker, S., & Bloom, P. (1990). Natural Language and Natural Selection. *Behavioral and Brain Sciences*, 13(4), 707–784.
- Prigogine, I., & Stengers, I. (1988). *Order out of chaos: man's new dialogue with nature*. London: Fontana Paperbacks.
- Riede, T., Tokuda, I. T., Munger, J. B., & Thomson, S. L. (2008). Mammalian laryngeal air sacs add variability to the vocal tract impedance: Physical and computational modeling. *Journal of the Acoustical Society of America*, 124(1), 634–647.
- Savage-Rumbaugh, E. S., McDonald, K., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General*, 115(3), 211.
- Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., & Bates, E. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, i–252.
- Schleicher, A. (1869). *Darwinism tested by the science of language*. London: John Camden Hotten.

- Schwartz, J. H., & Tattersall, I. (2000). The human chin revisited: what is it and who has it? *Journal of Human Evolution*, *38*(3), 367–409.
- Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*, 411–417.
- Smith, A. D. M. (2014). Models of language evolution and change. *WIREs Cognitive Science*, *5*, 281–293. <https://doi.org/10.1002/wcs.1285>
- Smith, K. (2011). Learning Bias, Cultural Evolution of Language, and the Biological Evolution of the Language Faculty. *Human Biology*, *83*(2), 261–278.
- Smith, Kenny, Brighton, H., & Kirby, S. (2003). Complex Systems in Language Evolution: The Cultural Emergence of Compositional Structure. *Advances in Complex Systems*, *6*(4), 1–22.
- Steels, L. (2000). Language as a Complex Adaptive System. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. Merelo, & H.-P. Schwefel (Eds.), *Parallel Problem Solving from Nature, PPSN VI* (pp. 17–26). Berlin: Springer.
- Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535.
- Tomasello, M. (1995). Language is Not an Instinct. *Cognitive Development*, *10*, 131–156.
- van der Ham, S., & de Boer, B. (2015). Cognitive bias for learning speech sounds from a continuous signal space seems non-linguistic. *I-Perception*, *6*(5), 1–8.

- van Kampen, N. G. (1992). *Stochastic Processes in Chemistry and Physics*. Amsterdam: Elsevier Science B. V.
- van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., ... Merrill, M. (2003). Orangutan cultures and the evolution of material culture. *Science*, *299*, 102–105.
- Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., ... Slocombe, K. E. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, *25*(4), 495–499.
- Whiten, A. (2017). Social learning and culture in child and chimpanzee. *Annual Review of Psychology*, *68*, 129–154.
- Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., ... Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*(6737), 682.
- Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, *50*(1), 56–64.
- Zuidema, W. (2005). *The major transitions in the evolution of language* (PhD Thesis). University of Edinburgh, Theoretical and applied linguistics, Edinburgh.
- Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, *37*(2), 125–144.