Integration Hypothesis: A Parallel Model of Language Development in Evolution*

Shigeru Miyagawa
MIT/UTokyo

Abstract

There are generally two views of how language emerged in evolution: emergent and gradual. The emergent view holds that language appeared relatively rapidly within the last 100,000 years, possibly due to some minor mutation. The gradualist view postulates stages of "protolanguage" that began as a simple system that progressively developed into ever-complex systems until language as we know it emerged. The original protolanguage may have been singing, as Darwin conjectured, or lexical in nature as proposed by a number of linguists. Human language is enormously rich and complex, which makes it difficult to imagine that all the components of it emerged somehow out of the blue in recent evolutionary time, yet there is no evidence for such a system earlier in evolution. The Integration Hypothesis holds that language is an integration of two independently occurring systems in nature that underlie communication. One system, exemplified by the alarm calls of primates, is the Lexical system, which is composed of isolated units of utterance that typically have a specific referent, such as "leopard," "snake," and "eagle" we see in the calls of Vervet monkeys. The Expression system, associated with birdsong, creates patterns without use of lexical items. Each system has developed over a long span of time, millions, and possibly hundreds of millions of years. At some point in recent evolutionary time, the two systems, L and E, integrated uniquely in humans to give rise to language, which gives the appearance of rapid emergence. I will speculate on how the integration may have been triggered.

1. Introduction

Language is a defining trait of who we are as human beings. We use it constantly and in many ways it is fundamental to our unique successes as a species. How did human language emerge in evolution? Our best estimation is that it appeared within the past 100,000 years, around the time that our ancestors were migrating out of Africa into the Eurasian continent. Because language doesn't fossilize, we have no direct evidence for what precisely happened to give rise to language. One might wonder if it is even worth asking the question about language in evolution if we cannot come up with direct proof. Indeed, back in 1866, the Linguistic Society of Paris banned all debate about evolution of

*I am grateful to Esther Clarke, Takashi Morita, Vitor Nóbrega, Kaz Okanoya, and Bridget Samuels for numerous suggestions on an earlier version of this chapter.

language because of lack of empirical evidence. It isn't suitable for serious study, they said. But we are in a very different world today, a world in which we know a great deal more about our brain and the brain of other animals, primate communication, birdsong, and, most importantly, human language, some or all of which will form a promising foundation for exploring the question of how language emerged in evolution.

The great amount of knowledge we have amassed over the past 150 years could help to solve many of the puzzles that form the mystery behind the emergence of language. But that is not the only reason why we should pursue this question. The emergence of language is one of the great mysteries of evolution and, some say, one of the hardest problems in science. This great mystery about language and the mind that produces it is all the more perplexing when we note there is an enormous gap between our closest relatives, primates like the chimpanzees and the bonobos, and humans in terms of our cognitive abilities. Our closest relatives don't produce art, they don't cook with fire, they don't fight for ideals as opposed to just fighting common enemies, and, of course, they don't have language like we do. It is a trait unique to humans and not shared with other branches of the same monophyletic group (Tallerman and Gibson, 2009). Other members of the group do have modes of communication, but human language possesses richness and complexity far beyond anything else that we see in the animal kingdom. The mind that makes language possible is so much more powerful than what we see elsewhere in the animal world that Darwin himself commented, "the difference between the mind of the lowest man and that of the highest animal is immense" (Darwin 1871:100). In a similar vein, Alfred Russel Wallace, a leading evolutionary thinker who was a contemporary of Darwin, noted that evolution should have endowed the humans with a brain a little better than that of an ape, yet what we ended up with is far more powerful than what a gradual evolutionary adaption would predict.

The problem of how language emerged in evolution is complex, just as complex as language itself. Not surprisingly, just as there is debate about every aspect of human language including its core function — is it primarily for internally representing thought, or is it for communicating ideas and emotions externally, or is it for some other purpose? — there is debate about what evolutionary steps contributed to the development of human language. I will review some of the major proposals from Darwin (1871) to more recent proposals by linguists, particularly by Bickerton (1990, 2002, 2014) and Jackendoff (1999, 2002), and an extension of it by Progovac (2010, 2015). I will argue that there are promising components to these proposals, but also shortcomings. I will then present the Integration Hypothesis of language evolution (Miyagawa et al. 2013, 2014), which incorporates Darwin's idea that language began as song much like birdsong, and also the proposal by linguists that there was a phylogenetic precursor of language, commonly referred to as "protolanguage" (e.g., Hewes 1973). An important question that has been posed about language is, how quickly did it emerge? Was it gradual, traversing through possibly many stages of protolanguage before arriving at language as we know it today? This is the so-called gradualist view, and it has many prominent proponents (e.g., Bickerton, 1990, 1995, 1998, 2000, 2014; Pinker and Bloom, 1990; Newmayer, 1991, 1998; Pinker, 1994; Jackendoff, 1999, 2002; Tallerman, 2007; Hurford, 2012; Progovac, 2006, 2008, 2009, 2010, 2012, 2015; Progovac and Locke, 2009). Or did the essence of today's language form rapidly? This is the so-called emergent view (Berwick, 1998; Hauser et al., 2002; Chomsky, 2005, 2008, 2010, 2012, 2013, 2014; Berwick and

Chomsky, 2011; Di Sciullo, 2011, 2013, 2014; Bolhuis et al., 2014; Miyagawa et al. 2013, 2014; Nóbrega and Miyagawa 2015). The rapid-development hypothesis comes in two versions, the better known being that something happened, possibly a mutation, that led to a rapid emergence of language where nothing like it existed before (e.g., Berwick and Chomsky 2011). The other view is that language arose rapidly from integrating pre-adapted systems that existed independently in nature (Miyagawa et al. 2013, 2014). The latter is the Integration Hypothesis, which I will take up to show its advantages and also a shortcoming, a shortcoming shared with the gradualist view proposed by Bickerton, Jackendoff, and Progovac. I will speculate on a possible solution to the problem, which will allow us to utilize Darwin's "song" idea and combine it with the proposals by linguists that are based on a sophisticated knowledge of human language.

2. Did protolanguage exist?

Homo sapiens emerged in Africa some 200,000 years ago, and their brain continued to evolve. It was perhaps some 60,000 - 100,000 years ago that a number of things appeared that demonstrate enhanced cognitive capability, including refined tools, carved and painted art, and sophisticated weapons. It isn't clear whether language emerged concurrent with these other achievements of high-order cognitive capability, but certainly language requires immense cognitive capability that is consistent with these other achievements of the *Homo sapiens*. Another piece of evidence for when language emerged is the discovery of so-called ochres in the Blombos Caves in South Africa. Ochre is an iron-rich mineral, and more than 8,000 pieces of ochre-like material have been found in the Blombos Caves, and they have been dated back to 75,000 - 100,000 years ago. Some, like this one, have engravings and incisions.

Figure 1. Ochres from the Blombos Caves (Evolutionary Studies Institute University of the Witwatersrand)

Though some have questioned it, Tattersall (2009) and others have argued that these patterns represent early abstract or symbolic depiction, possibly similar to language. It is certainly within the approximate time span when it is believed that human language arose.

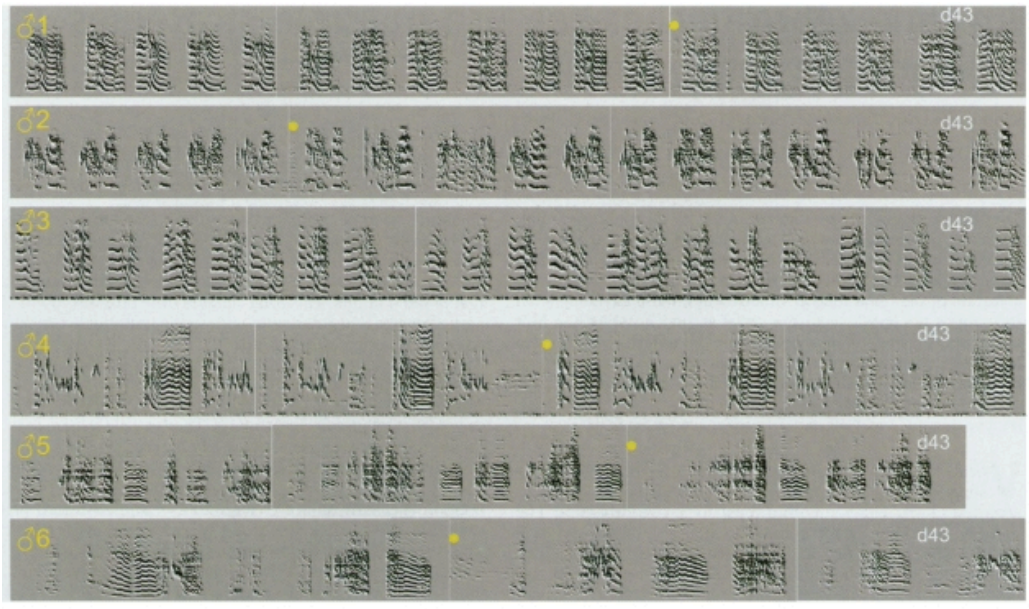
Although there is disagreement on just about every aspect of human language, one point that would be difficult to deny is that it is an enormously complex system composed of a countless number of elements — sounds, words and subwords, syntax, and meaning. Given this complexity, it would be natural to posit that human language must have developed gradually from an earlier, simpler protolanguage system. I will briefly review two proposals for protolanguage, the song protolanguage of Darwin (1859,

1871) and the lexical protolanguage of Bickerton, extended by Jackendoff. For other proposals, see Fitch (2010), who gives an excellent summary and critique of the major protolanguage proposals. For the development of cognition in Homo sapiens, see, for example, Mithen (2014).

2.1 Song Protolanguage

Darwin (1859, 1871) was the first to note that there may be a connection between birdsong and language; he described it as the "nearest analogy to language." In the same way that humans have an instinct to speak, songbirds have an instinct to sing, and just as language is learned, so are birdsongs learned, in both cases the learning process opening up the possibility of variety, as in dialects in both human language and birdsong. In both language and birdsong, there is a stage of acquisition that precedes mastery; in humans this stage is infant babbling and in juvenile birds it is the singing of "subsongs." On the last point of acquisition, recent studies show a surprising parallel between language and birdsong beyond simply sharing a pre-mastery stage. A birdsong, such as that of zebra finch, is composed of notes that are combined to form syllables, and syllables combined to form motifs, which are then combined to form a complete song. Liu, Gardner, and Nottebohm (2004) identified two learning styles, the “serial repetition” strategy and the “motif" strategy. For a juvenile bird that adopts the serial repetition strategy, an approximation to one syllable of the model is repeated many times, and because the unit of repetition is small, each syllable is clearly articulated. In the motif strategy, the juvenile bird adopts a global imitation of its father's song (only males sing), and because the unit of repetition is larger, the articulation is noisy and imprecise. The following spectrograms show the striking difference in these two strategies among two groups of 41-day old zebra finch juvenile males.

Figure 2 Serial repetition and motif strategies of song learning (Liu, Gardner, Nottenbohm 2004)



In the upper half, we see three spectrograms of juveniles that adopt the serial repetition strategy, with each syllable clearly sung. In the bottom half are spectrograms of juveniles with motif strategy; given the larger unit being sung, the overall song lacks the precision of the other strategy. Regardless of the strategy that a juvenile adopts, by the time it matures, it is able to sing the song perfectly. As it turns out, the two strategies for acquisition are also found in human language acquisition. O'Grady (2005) (see also Lipkind et al. 2013) points out that a human infant may adopt one of two styles of learning: the "analytic" style, which focuses on breaking speech into its smallest component parts to produce short, clearly articulated, one-word utterances in the early stages; and the "*gestalt*" style, in which the infant memorizes and produces relatively large chunks of speech that are often poorly articulated that correspond to entire sequences of words in the adult language. Just as with songbirds, the human infant, regardless of the style of acquisition it adopts, ultimately acquires the language perfectly.

2.2 Lexical Protolanguage

One idea that a number of linguists have suggested is that before human language became the full-fledged system that we are familiar with, it went through a simpler linguistic stage, what Fitch (2010) calls "lexical" protolanguage. According to Bickerton (1990, 2002, 2014), this earlier stage was composed typically of a simple utterance that had no structure. One way to think about this, as Jackendoff (1999) suggests, is that protolanguage was today's language minus syntax, a point also suggested by Bickerton (1990). So, it is just made up of words only. In other words, for linguists such as Bickerton and Jackendoff, protolanguage was the capacity to use unstructured symbolic units. How did such a system arise, if, indeed, it did exist? Bickerton suggests that one possibility for the emergence of such a system of communication is to convey the location of a dead prey in scavenging. This was a time when our ancestors were hunter-gatherers. Why would our scavenging ancestors require such a system? It is surmised that this system allowed the hunter-gatherers to communicate the location of dead prey at a distance. To be successful in claiming the dead prey from competitors, these pre-humans had to recruit members of their community in large number and communicate to these members the location of the prey and some idea of its size and the desirability of being able to claim it. This requires displacement, which is the ability to communicate about things that are not in the immediate vicinity of the speaker-hearer. Human language is adept at displacement; so are the systems employed by honeybees and ants for similar resource-locating purposes. The idea is that, in humans, protolanguage emerged to make displacement possible. This is an intriguing idea. Did language actually go through such a protolanguage stage? It is hard to say, of course, because it happened so long ago. We will return to both the song and lexical protolanguages later in the chapter.

3. *Both* Gradual and Rapid

The gradualist viewpoint based on protolanguage makes sense if one were only focused on language and how it may have evolved over time. Given the complexity of

human language, it would be reasonable to assume that this complex system began as a simpler mechanism that over a span of evolutionary time came to take on the kind of complexity and richness we are familiar with. But there is a crucial point here: evolution is about living organisms, but language is not a living organism. From this perspective, languages may change, but they don't evolve (Chomsky 2014). But clearly something happened in evolution to give rise to language. Was it an event, or series of events, that took place gradually over a long span of evolutionary time, or was it fairly rapid in nature? My suggestion is that it was *both*. I begin with the gradualist part of my idea, leaving the rapid part until the next section.

The time span I have in mind for the gradual development needed for the eventual emergence of language is much longer than the evolutionary time implied in the various models of protolanguage. In fact, it is millions of years, and in one case, hundreds of millions of years. It is our brain that provides the cognitive capacity to produce language, and the brain we have inherited has been growing ever larger long before *Homo sapiens* appeared in Africa 200,000 years ago. *Homo habilis*, which appeared on the scene some two million years ago, started to show an increase in brain size (Deacon 1997). Why is this? Some attribute the growth in brain size to eating habits. The two systems, or organs, that consume the most energy in our body are the digestive system and the brain. As our ancestors became adept at preparing food, including grinding it, and most importantly, heating it with fire to soften it, less energy was needed for digestion, and the brain received the benefit of the excess energy and grew larger.

The *Homo sapien* brain is not just larger relative to the body size, but well developed in regions that contribute to higher order cognitive functions such as the prefrontal cortex which is responsible for high-level cognitive activities.[1] Language requires a number of special components in the brain and in our physiology for speech. Within the brain, two of several important centers for language are Broca's area and Wernicke's area. Both are located almost always on the left hemisphere. Broca's area, located in the frontal lobe, supports speech production, while Wernicke's area, in the superior temporal gyrus, supports speech perception. Both are needed for language. So the question is, when did these regions of the brain develop? Did they develop around the time that language arose in evolution, 100, 000 years or so ago? There is evidence from study of skulls of *Homo erectus* that these centers may have been around in the brains of our ancestors over a million years ago, long before language actually emerged (Wynn 1998). In addition, there is evidence that homologues of these structures are present in some great ape brains today (Gannon et al. 1998; Cantalupo and Hopkins 2001) suggesting they may have developed over 9 million years ago.

This idea that some of the key centers of the brain that support language have existed for several million years, maybe more, is part of a picture that is emerging recently about who we are. Although it was only recently in evolutionary time that we began to see evidence of high cognitive achievements, such as painted art, sophisticated weapons, refined tools, and language, the brain circuits that make them possible apparently have been present much longer. In recent articles, it is pointed out that the human cortex, where higher-order cognitive computation is carried out, has similar cell types, patterns

[1] Recent studies point more on expansion of the temporo-parietal junction. (e.g, Bruner, E., & Iriki, A. (2016). Extending mind, visuospatial integration, and the evolution of the parietal lobes in the human genus. Quaternary International, 405, 98-110.

of wiring, and gene expression as other mammals (Calabrese and Woolley 2015, Harris 2015). Furthermore, the computations that these brain circuits make possible also occur in birds. It is as if there are only a limited number of brain microcircuit designs that nature came up with and these circuits have been repurposed to make the higher order cognitive achievements possible in humans.

Among the recent studies, one in particular has direct implication for Darwin's song protolanguage in which he surmised that pre-humans sang and this ultimately led to language. Pfenning et al. (2014) look at vocal learning as reflected in regions of the brain of songbirds (zebra finch) and humans, and also birds that don't sing (dove, quail), and a primate that doesn't have language (macaque). Taking a computational approach, they screened gene expression databases from humans and vocal-learning birds as well as the non-vocal-learning birds and primate. Their study indicates molecularly analogous regions that are homologous for song and speech for vocal-learning birds and humans. They also demonstrate that non-vocal learning birds and macaques do not share in any significant way these traits. The idea that humans and songbirds have homologous regions that are not found in non-singing birds and also in macaques raises an interesting question. Are these shared regions due to convergent evolution or to a common progenitor? Convergent evolutions are sometimes known to occur among unrelated living beings, as similar solutions may arise for similar problems (Gould 1976), as in the example of the emergence of the eye in unrelated organisms (Ogura et al. 2004, Fernald 2006). On the convergent evolution view, the shared regions are analogies of each other. The other possibility is that the regions with similar functions in vocal-learning birds and in humans descended from a common ancestor. Given that the ancestors of present-day birds and mammals split 300 million years ago (Benton 1990), this suggests that, on the common progenitor view, the genetic resources for singing in primates existed over 300 million years ago. Darwin probably did not imagine that the song protolanguage stage lasted that long, and our ancestors may not have sung for all of that time, but just as we saw for Broca's and Wernicke's areas, what we see is that the genetic resources may have been present long before language emerged. On the convergent view, we are looking at a time depth of at least around nine million years if the existence of regions similar to the Broca's and Wernicke's areas in the brain of some great apes is an indication of the age of these regions. On the common progenitor view, the brain centers for songbirds imply a much older evolutionary history. But what does singing have to do with human language? While birds still sing today, we see on the primate side only four species that sing: titi monkeys, indris, tarsiers and gibbons. Of these four, gibbons are the only apes (Hylobatidae) and thus the most closely related to humans. Their songs have been studied for decades (Marshall and Marshall 1976, Haimoff 1984) but have only recently been compared to human language (Clarke et al. 2006). I will present a hypothesis for how language emerged in evolution that considers the system underlying singing to play a critical role, thus inheriting Darwin's idea of a singing pre-human, and linking it to the regions of the brain for singing that may have existed for million years and perhaps for over 300 million years.[2]

[2]Samuels (2015) argues that songbirds have capacity to generate sound structure analogous to human language phonology.

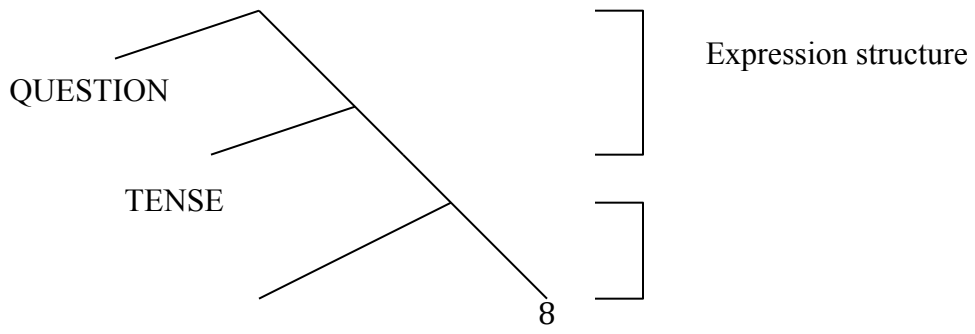
4. Basics of Human Language

According to the UNESCO report, *Endangered Languages*, there are approximately 6,000 languages in the world. Some may question this number as an underestimation — for example, just in Papua New Guinea we find 820 languages, and there are many other regions of the world that hold a large number of languages. The precise number is hard to pin down because sometimes it is difficult distinguish between dialects and languages. However many languages there are, one thing is certain: every language has the function, as Darwin described it, to connect "definite sounds with definite ideas" (Darwin 1871: 54). How does language do this? In the linguistics literature, the proposal is made that each sentence is composed of two layers of meaning, called "duality of semantics" (Chomsky 2005). We can use the following example to demonstrate this duality.

(1) Did you eat bread?

This sentence contains three content words, *you*, *eat*, and *bread*. These words may occur in a variety of contexts with a fixed meaning, such as *I want you to eat bread, you shouldn't eat bread, You seem to eat bread everyday.* This is the *lexical* layer, containing the meaning of individual words of a sentence. The sentence also contains the word *did*, a function word that has not just one, but two functions. By inflecting for tense, it indicates that the event represented in the sentence took place in the past, and by occurring at the head of the sentence, it signals that the sentence is interrogative in form. Tense and interrogative are two elements in the other layer of meaning, which I call the *expression* layer (Miyagawa 2010). The expression layer constitutes a chunk of the expression, typically a sentence, and its purpose is to give shape to the expression, such as interrogative, and any meaning that is associated with the entire expression, such as tense. This is in sharp contrast to the *lexical* layer, in which the meaning is represented at strictly the local level of each individual word. The two layers of meaning differ in two other significant ways. The lexical layer is composed of content words, and an adult English speaker knows around 60,000 such words. In contrast, the expression layer is composed of function words, and these number in the single digits to less than fifteen depending on the theory one adopts. Also, the words of the lexical layer connects to specific meaning, often, though not always, referring to an entity in the real word, such as *table, school, horse*. The expression layer provides the shape of the utterance, and it commonly communicates the intention of the speaker, such as the intention to ask a question, make a statement, issue a command, and so forth (Austin 1962). The two layers of meaning are represented as follows for the earlier example.

(2) Duality of Semantics (Chomsky 2005, Miyagawa 2010)



Lexical structure

you eat bread

QUESTION indicates that the entire expression has the form of interrogative, and represents the speaker's intention to ask a question, while TENSE contributes to the overall expression by indicating the tense of the utterance. Each individual word in the lexical layer is represented in the lower structure.

In Miyagawa et al. (2013), I argued, based on the work of a number of linguists, that although the expression and lexical layers are intricately intertwined in human language, they can be viewed as separate components each with its own specific properties, as we saw. The two components are in fact so fundamentally different that it appears that they just happen to occur together in language, but they could just as well function as separate systems. In Miyagawa et. al (2013, 2014), I in fact propose that the two layers correspond to two distinct systems that occur independently in nature. I call it the Integration Hypothesis of language evolution because the two layers happened to get integrated uniquely in humans and gave rise to language as we know it today. By separating language into these two components, we are able to entertain a range of possibilities that are difficult to consider in the other proposals for language in evolution. Given the richness and complexity of human language, it is certainly challenging to imagine that it developed rapidly within the last 100,000 years, yet there is no evidence of language prior to that time, as far as we can see. How do we make sense of this enormously complex system that just "popped up" 100,000 years ago? By the Integration Hypothesis, the process of integration of the two independent systems occurred in recent evolutionary time. But the two systems themselves are older, much older, their progenitors being millions, or in one case, possibly hundreds of million years old. Below, I turn to the Integration Hypothesis.

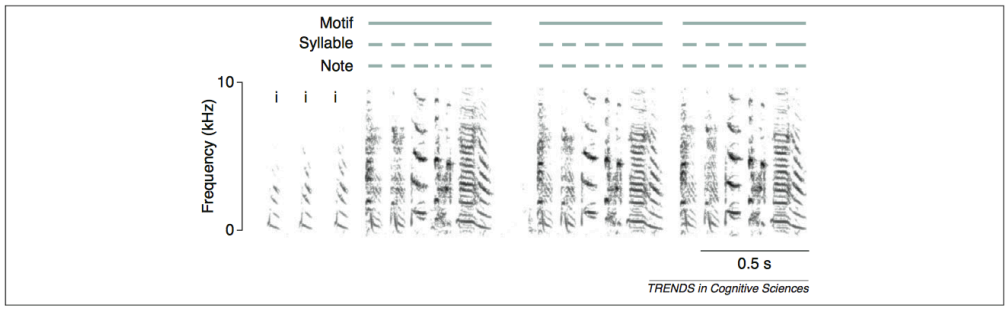
5. Lexical and Expression layers: Separate systems in nature

If the two layers of language, *lexical* and *expression*, have correlates in independent systems in nature, what are these systems? The calls of monkeys and apes are a natural candidate for the *lexical* system. There is a large body of literature on this topic (Seed and Tomasello 2010), an earlier work of which is on the Kenyan Vervet monkeys (Seyfarth et al. 1980), which possess alarm calls for pythons, eagles, and leopards. Sticking to these simplest lexically based systems, what we see is an uttered object that correlates with a particular real-world state of affairs. An important point about this system is that each uttered object is co-extensive with a real-world state ("holistic" in the sense of Wray 1998). A simple, and possibly a simple-minded, way to view this system is that it is a collection of vocal gestures that have a specific referent in the real world (Miyagawa et al. 2013). It is a "pure" L(exical) system in that each unit is an isolated entity with a definite "meaning." The lexical protolanguage view points to this L system in our

ancestors as forming the progenitors to language.[3]

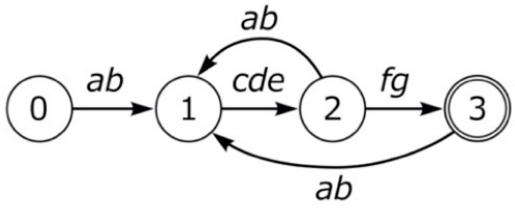
Miyagawa et al. (2013, 2014) links the E(xpression) layer of human language to the system underlying birdsong.[4] As Darwin first noted, there is a list of striking resemblances between birdsong and language, a list to which we can also add: a need for external input; sensitive developmental periods ending at sexual maturity; hemispheric lateralization, and motor-auditory rehearsal systems (Bolhuis et al. 2010). There is one striking difference: lexical items in the sense of human language are absent from every variety of birdsong that has been studied, so that a birdsong appears to be a phonological syntax without a lexicon (Marler 2000). Every birdsong is a system that generates particular patterns, as in the case of the well-studied song of the zebra finch, which has a highly restricted set of "notes" that combine to form syllables, and syllables into motifs, and motifs into a complete song "bouts" (Berwick et al. 2011).

Figure 3 The song of zebra finch (Berwick et al. 2011)



The zebra finch song is one of the simplest; other vocal learning bird species possess more complex patterns with branches, loops, and repetitions (Berwick et. al 2011).

Figure 4 Bengalese finch (Okanoya 2004)



A unique property of human language is that a regular grammar (type 3) is inadequate for modeling it (Chomsky 1956). But this is not the case for the two proposed antecedents of human language. The lexical systems that employ isolated uttered units that correlate with real-world references, such as the alarm calls of Vervet monkeys, are clearly couched in a simple regular grammar. The other layer, the expression layer, finds

[3]Although I consider the L system to consist of independent vocal gestures, there are studies that suggest that primates can use multiple calls to construe novel meaning (Dessalles 2007, Arnold and Zuberbuhler 2006, Tallerman and Gibson 2011).

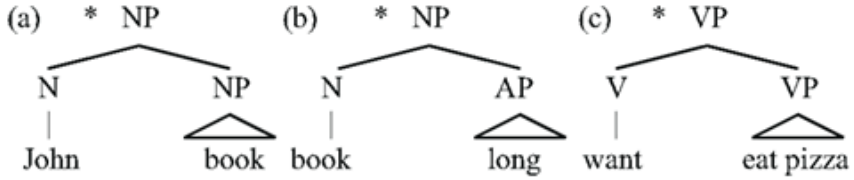
[4] See Okanoya (2002) for an earlier proposal that links birdsong syntax and human language.

its antecedent in the kind of system underlying birdsong. Birdsongs have specific patterns, and these patterns can be complex, as in the example of the Bengalese finch, which loops back to various positions in the song and lead to considerable variation. Nevertheless, all known birdsongs can be described as a *k-reversible* finite state automaton, a form of regular grammar (Berwick et al. 2011); for example, we do not see center-embedding, which was an example Chomsky used to prove the inadequacy of regular grammar for describing human language. While suggestions have been made that certain bird species can acquire recursive syntactic structures reminiscent of human language either through conditioning (Gentner et al. 2006) or spontaneously (Abe and Watanabe 2011), these results remain controversial (Beckers et al. 2012) and unconfirmed.

The Integration Hypothesis proposes that these two major systems in nature that underlie communication, L(exical) and E(xpression), integrated uniquely in humans to give rise to language. One challenge to the Integration Hypothesis is that, as noted above, contemporary languages are known to be based on a system more powerful than a regular grammar. As evidence for the Integration Hypothesis, in Miyagawa et al. (2013, 2014), we demonstrate that when we consider the two components of contemporary languages separately, L and E, each component may be characterized by a regular grammar, thus reflecting their antecedents in nature. It is only when the two systems integrated that gave rise to a system more powerful than what we see elsewhere in the animal kingdom.

Let us begin with the L layer. A trait of L units is that they don't combine directly, just as the discrete units of the Vervet alarm calls do not combine to form new calls (Miyagawa et al. 2013).

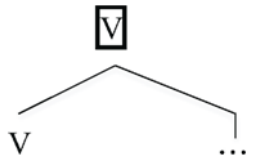
(3) Impossible lexical structures



To make these combinations possible, something from the E layer must intervene: the D(eterminer) represented by apostrophe *'s* (*John's book*), the copula that carries tense, an E layer phenomenon (*book is long*), and *to*, which helps to form the E layer of a clause (*I want to eat pizza*). Later, we will look at compounds, which ostensibly pose a challenge to this view that L units do not directly combine.

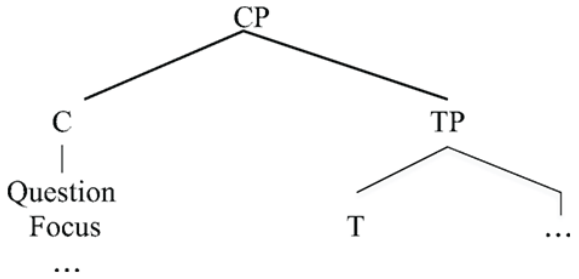
To see the regular-grammar status of the E layer of human language, let us begin with a unique feature of human language not seen in other systems in the animal world, the feature commonly called "labeling" (Chomsky 1995). Given a word, its lexical category (Noun, Verb, etc.) forms the "label" of the larger phrase that contains it. For example, for the pair *eat* and *bread*, the verb *eat* labels the larger phrase, *eat bread*, as a V(erb), forming a Verb Phrase.

(4) Labeling



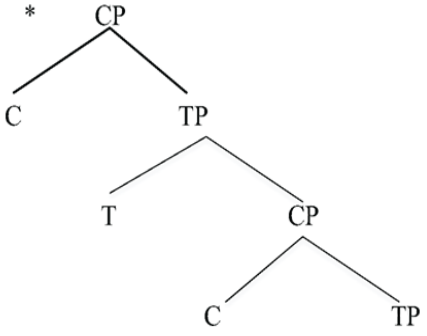
This property of labeling gives human language its unique ability to form hierarchical structures (Chomsky 1995, 2008; Hornstein 2009) and gives human syntactic structure the property of "discrete infinity" (Chomsky 2000) through recursively merging and labeling structures. However, as noted in Miyagawa et al. (2013), there is a severe limitation on the depth of the hierarchy when we isolate structures in the E layer. As we saw earlier, the E layer can contain Tense; there is a second item, conventionally labeled "C(omplementizer)" that contains a range of expressive phrases such as Q(uestion), F(ocus), and so forth.

(5) Expression structure



C and T are the two most frequently cited labels within the E layer. Strikingly, these labels cannot occur as hierarchical structures of arbitrary depth; rather the CP-TP structure can only be one layer deep, a limitation consistent with the E layer being characterizable as a regular grammar.

(6) An impossible E-layer structure



The limitation is that the E-layer is restricted to a depth-one hierarchical structure, similar

to the systems we see in the Bengalese finch and nightingale songs.[5] This suggests that the E layer of human language closely reflects the birdsong structure, a system of regular grammar. While there are theories of linguistics that posit multi-layer within E (Rizzi 1997), there are alternatives that do not assume such a multi-layer (e.g., Miyagawa 2010).[6]

6. Progenitors of human language

The previous proposals about how human language developed all have some "beginning" that is a simpler stage of today's language, and through a progression of stages, each more complex than the earlier one, we arrive at today's human language. For example, and as already mentioned, lexical protolanguage, proposed by linguists (e.g., Bickerton 1990, 2000, 2014; Jackendoff 1999, 2002), views human language as having undergone a sequence of stages, from a one-word stage to a more complex system of expression; in the case of Jackendoff, and more recently Progovac (Progovac and Locke 2009, Progovac 2012, 2015), this second stage is characterized by the combination of single words into compounds through a rudimentary recursive *n*-ary operation that generates flat structures (Progovac and Locke 2009, Progovac 2012, 2015). Once equipped with this proto-language that can form a two-word sequence, a primate's brain becomes ready at some point to create structures that ultimately lead to sentence formation. Why did syntax emerge? If it is the case that in protolanguage, any two words could be strung together, there would be no structure to define the relationship between the two words, so that there is a great deal of burden on the context to give meaning to the combination. For example, if one combines *child* and *picture,* as *child-picture*, we cannot tell if this means a picture drawn by a child, a picture that depicts a child, a picture for children, and so forth. One way to think about the reason why the protolanguage developed structure is to clearly mark the relationship between the combined items (Jackendoff 1999). This reduces burden on the brain to compute the meaning. From this, the system could have developed structures that define such relationships as modifier-modified as in *blue sky*, verb-complement, such as *eat bread*, and an agent of an act, such as *horse-rider*. Linguists such as Jackendoff and Progovac point to the existence of

---

[5]Arsenijevic and Hinzen (2012) also notice this limitation in the E layer and attribute this limitation to meaning. We believe that characterizing this limitation in terms of finite state grammar captures the restriction, which we believe holds of all kinds of E layers and not just the CP-TP structure, which is what Arsenijevic and Hinzen account for.

[6] Tallerman (2016) criticizes the idea that the expressive layer of human language parallels the structure of birdsong. She points out that while the elements of birdsongs always occur in an ordered sequence, in human language elements may undergo permutations — movement as called in linguistics — that alter the ordering sequence of items. This criticism is based on a fundamental misunderstanding of the expressive layer. As noted in Miyagawa et al. (2013, 2014), an operation such as movement is only possible after the two layers, E and L, integrated. The point is quite simple: what is moved are words in phrases, hence L as well as E must be involved. Tallerman incorrectly notes that movement is strictly handled by the E component. Rather, a non-finite state operation such as movement does not occur in E layer in isolation, but is the outcome of integration (Miyagawa et al. 2013, 2014).

certain compounds (*daredevil*) in contemporary languages as living fossils of an earlier, non-syntax stage of language.

Let us consider the lexical protolanguage proposal by looking at the most recent work on the topic — Progovac (2015). This work, by a scholar well-versed in modern linguistic theory and with the major literature on evolution, has amassed together considerable data from contemporary languages to argue that from its initial one-word stage, language evolved through progressively complex stages. A unique feature of this work is that Progovac specifically argues that each of these stages reflects the syntactic structure postulated in the modern linguistic theory called Minimalism (e.g., Chomsky 1995). Thus, from a one-word stage, language evolved into a two-word stage, and eventually it evolved to stages that progressively took on layers of functional structures, as in vP > TP > CP. Her interesting claim is that we find each of these stages as living fossils or in some related form in contemporary languages. Whether one agrees with her conclusion or not, the extensive data she presents is fascinating in and of themselves; I found particularly interesting the numerous "exocentric" compounds in Chapter 6 from Serbian and other languages. Independent of issues of language in evolution, these data, and other data in other chapters, present a body of empirical material for carrying out research on topics such as compounding and phrase structure. Progovac also presents the core of the recent linguistic theory (e.g., Chomsky 1995) in a way that is comprehensible to non-linguists, which is a valuable service to the field.

There are some issues with her work. Progovac considers the emergence of language as an evolutionary process, but as already noted, living entities evolve, and language is not a biological entity. It is telling that all the examples Progovac gives as examples of evolutionary change are biological in nature except language — lungfish, body hair, eye. So, if she is correct in her description of how language developed, one must wonder precisely what this development is telling us about the organism that produced each of these stages.

Furthermore, if we consider "one-word" and "two-word" systems, they are fundamentally different. In "one-word" systems, such as the Vervet alarm calls and the emotional interjections in contemporary human languages (*ouch*, *wow*), the units do not combine with other units to form a new unit. This is because each unit of utterance is co-extensive with the referent or the emotion it represents; since they don't have any lexical category, these single units are better called "roots" (Nóbrega and Miyagawa 2015). It is simply not possible to combine the alarm calls such as *leopard* and *eagle* because each has an independent reference and there is no sense in which one unit can combine with another unit to form a third unit with some other reference, unless these units undergo fundamental change. Same goes for human language interjections: there is no sense in which *ouch* and *wow* could combine to give a third emotional interjection simply because each interjection is a fully independent and isolated unit of utterance that expresses a specific emotional state. In contemporary languages, combining words into compounds or sentences minimally requires that each word ("root" to be more precise) be assigned a category (noun, verb, etc.), which automatically gives the word more structure than an unstructured "root" form of alarm calls or interjections; thus a word in a compound is complex (see Di Sciullo 2012, 2013, 2014).[7] As noted in Nóbrega and Miyagawa (2015),

[7]This is similar, and possibly the same, as the notion of "lexical envelop" that allows roots to merge with other "enveloped" units (Spelke 2003, Ott 2009, Boeckx 2011).

if one looks at compounds beyond those used as evidence for "living fossils," one finds structure that is just as complex as a full sentence, thereby questioning the idea that there was an unstructured two-word protolanguage stage. From this viewpoint, the four stages that Progovac suggests should be questioned (she refers to her proposal as a three-stage model, but I am including the "zero" stage as the first stage).

(7) Four stages of language evolution (Progovac 2015: 15)
(o) One word stage;
(i) Paratactic stage: conjoin two words without creating hierarchy or headedness;
(ii) Proto-coordination stage: emergence of "conjunction/linker" that provide "segmental glue to hold the utterance together" (p. 13), presumably making multiple-word utterances possible;
(iii) Specific-functional category stage: "specific functional categories become available" (p. 13). "it is only at this stage that hierarchical structure, Move, and recursion become available." (p. 14)

Based on what we have said, there are really only two crucial stages, the one-word stage, which is made up of isolated roots, and a stage that Progovac calls the "specific-functional category stage," where "functional categories" — what I am calling the E system — becomes available. Once the E layer becomes associated with the isolated units of the L layer, combining the L units under Merge is possible, with labeling of the newly created structure. From this perspective, there would be no need to postulate the paratactic and proto-coordination stages. That also makes the picture simpler for creating multi-word expressions. Instead of having to postulate "proto-Merge" which presumably joined roots, and which arguably does not occur in contemporary languages (Di Sciullo 2013, 2014),[8] we would simply need to postulate the same Merge that exits today that takes words associated with lexical category and other E properties.

Despite these issues, there are a number of points I agree with, including Progovac's suggestion that the progenitors of human language may have been around millions of years ago. I will return to this point below.

Turning to Darwin, a crucial feature of his view of how language emerged is that he separates cognitive abilities from the initial emergence of language-like behavior: "The mental powers in some early progenitor of man must have been more highly developed than in any existing ape, before even the most imperfect form of speech could have come into use" (Darwin 1871: 57). The protolanguage Darwin conjectured is musical in nature, and involved singing with the intention, same as songbirds and gibbons, of courtship and territoriality. This eventually developed into a full-fledge language, according to him, through imitations aided by signs and gestures that became associated with definite meaning.[9] Darwin goes on to add that the emergence of language aided in the further development of the mind, very much in the spirit of the German philosopher and linguist Wilhelm von Humboldt, whose work Darwin studied closely.

The idea that pre-humans sang appears to be particularly plausible given the recent genetic finding that humans with their speech, and birds with their song, share regions

[8]See Boeckx (2015) for an alternate view.
[9]See Jesperson (1922) for a different model of how song developed into language.

that are homologous for speech and song (Pfenning 2014). Where Darwin's conjecture becomes questionable is, how could a musical system, which lacks lexical units, come to take on such units built into the song?

The Integration Hypothesis differs from all other hypotheses of language development in that it does not postulate a singular "proto" system that developed progressively into ever-complex systems that ultimately ended up as language as we know it today. Any "proto" system runs into the kinds of problems I noted above. If one begins with a word (or more accurately, root, which has no structure), how does syntax develop out of it to make it possible to combine these words into two-word utterances and beyond, to sentences? If one starts with a musical form of protolanguage, how do words get introduced into this system? What the Integration Hypothesis suggests is that the sources for words and syntax existed independently as the L system underlying alarm calls and similar modes of communication, and the E system underlying birdsong (and possibly gibbon song). It is reasonable to assume that these two systems have been around for a long time, millions of years or even longer. At some point in recent evolutionary time, the two systems integrated uniquely in humans to give rise to a system that generates patterns as an E system, and contains lexical items as an L system. Why this happened is not clear, but we could imagine some things that had to have happened to make this integration possible. As noted earlier, there is a fundamental difference between one-word and two-word systems. While a one-word utterance can be co-extensive with the referent or an emotion, a two-word (or more) combination requires each word to be non-referential -- underspecified for meaning, in other words -- so that the two can combine to form a novel utterance that has its own unique reference. This means that each word must have undergone some process of abstraction. Thus, in *child-picture*, *child* does not refer to a definite entity, but it is a *kind* of an entity that shares essential properties (human, youth, etc.). The same goes for *picture*. What could have shifted the isolated utterances of the L system into these abstract entities that can be combined into patterns furnished by the E system? It is not clear, but again, we can see what must have occurred: each item, as a process of abstraction, took on category, such as noun, verb, and so forth. As a noun, for example, it no longer referred to a specific entity, but stood for a *kind* of an entity (*bread*). Furthermore, by taking on categorical identity, it is able to merge with another unit, and, crucially, to label the newly formed unit (V label for {eat bread}).

What about the E system? Birdsong is a pure E system that generates finite-state patterns without any specific reference to entities in the real world. It is, as already noted, a phonological syntax without a lexicon (Marler 2000). Just as some change must have taken place to the L system, from its "one-word" units that are co-extensive with the meaning they embody, the E system must have undergone some change to accommodate the members of the L system in order for integration to have taken place. What could that change be? One possibility is found in Darwin's proposal of musical protolanguage. He conjectures that for the singing of pre-humans to have transitioned to articulate language, they began to incorporate imitations of natural sounds that represented some entity, aided by gestures (Fitch 2010: 472), something also suggested by Farrar (1870) and Müller (1861). One way to think about this change is that the E system changed from a system of nonreferential patterns to one in which the patterns contained placeholders for items that refer to the real world. These placeholders are represented by imitation sounds and other

"referential" units. It then becomes natural to start inserting the members of the L system, which themselves have undergone change to be underspecified for meaning. As these members of the L system began to combine within the E system, they start to build structure in the way that lexical protolanguage proponents such as Jackendoff have conjectured. One crucial difference is that the building blocks of the structure are furnished by the E system, as opposed to the proposal of the protolanguage proponents, who conjecture that the structures somehow arose directly from L combinations.

Once the integrated system achieves this state of having L units that have category and an E system that can merge these L units into structures that can be labeled, we have the essence of the human language at work. The idea that human language emerged within the past 100,000 years does not mean that all the pieces of language developed within this time frame. All the essential pieces have existed for millions of years and developed gradually over a long evolutionary time span, and what happened to give rise to language was that these independently developed pieces integrated uniquely in humans to give rise to the kind of rich and complex system we know to be language. This gives the appearance of language having rapidly developed in recent evolutionary time, and while it is true that there is no evidence for language earlier in evolutionary history, we have ample evidence for the existence of L and E systems long before language emerged. It is the adventitious integration of these two highly-developed, pre-adapted systems that gave rise to the remarkable system of human language.

Among all the necessary conditions for integration, it appears that abstraction (or underspecification) of the L units is one crucial missing link. Other pieces were essentially there as independent systems: the units of L, and the pattern-generating capacity of E. For the units of L to combine by tapping the pattern-generating E system, the L units must undergo abstraction, possibly by being assigned a category such as noun and verb. It is not clear how this happened. One possibility, and I only mention it as a conjecture, is the notion of displacement that Bickerton incorporates into his proposal of protolanguage. Note that displacement requires abstraction. A unit of L that refers, for example, to leopard, can do so if the referent is present in the physical proximity. For an organism to convey the existence of an entity in the absence of this entity, the organism must convey the *idea* of the entity. By necessity, this involves some ability to conceptualize the *kind* of the entity involved, hence abstraction, which is required for integration. Even if this turns out to be true, how this abstraction leads to the uttered unit becoming associated with a lexical category remains a mystery.

7. Concluding remarks

Did language emerge rapidly or gradually? Given the rich and complex nature of language, it surely makes sense that the key components of language took a long evolutionary time to emerge. Yet, there is no evidence for such gradual development, in which language first appeared, say, as a one-word system followed by a two-word system and so forth. The Integration Hypothesis suggests a parallel development of independent systems in nature that underlie communication, the E and L systems, which allows us to capture the essence of both the emergent and gradualist views of language in evolution. Each system developed over a long span of time, possibly as long as three hundred

million years. At some point in recent evolutionary history, the two systems integrated uniquely in *Homo sapiens* to give rise to the kind of system we see today as language.

References

Abe, K. and Watanabe, D. (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat Neurosci*. Jun 26;14(8):1067-74. doi: 10.1038/nn.2869.

Arnold, K., and Zuberbühler, K. (2006) Language evolution: Semantic combinations in primate calls. *Nature* 441**,** 303.

Arsenijevic, B., and Hinzen, W. (2012). On the absence of X-within-X recursion in human grammar. Linguistic Inquiry Summer, 43(3), 423–440.

Austin, JL (1962) *How to do things with words*. Cambridge: Harvard University Press.

Beckers, G. J., Bolhuis, J. J., Okanoya, K., and Berwick, R. C. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport* 23, 139–145. doi: 10.1097/WNR.0b013e32834f1765

Berwick, R. (1998) Language evolution and the minimalist program: The origins of syntax," in *Approaches to the Evolution of Language: Social and Cognitive Bases*, eds. J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Cambridge: Cambridge University Press), 320-340.

Berwick, R.C., Okanoya, K., Beckers, G.J.L., and Bolhuis, J.J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences* 15**,** 113-121.

Berwick, R. C. and Chomsky, N. (2011). "The Biolinguistic Program: The Current State of its Development," in *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of Human Language Faculty*

Bickerton, D. (1990). *Species and Language*. Chicago, IL: Chicago University Press.

Bickerton, D. (1995). *Language and human behavior*. London: University College London Press.

Bickerton, D. (1998). "Catastrophic evolution: The case for a single step from protolanguage to full human language," in *Approaches to the Evolution of Language: Social and Cognitive Bases*, eds. J. R. Hurford, M. S. Kennedy and C. Knight (Cambridge: Cambridge University Press), 341-358.

Bickerton, D. (2000). "How protolanguage became language," in *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*, eds. C. Knight, M. Studdert-Kennedy, and J. R., Hurford (Cambridge: Cambridge University Press), 264-284.

Bickerton, D. (2014). *More than nature needs: Language, Mind and Evolution*. Cambridge, MA: Harvard University Press.

Boeckx, C. (2011) Some reflections on Darwin's Problem in the context of Cartesian biolinguistics, eds. Di Sciullo, A-M, Boeckx, C., *The biolinguistic enterprise: new perspectives on the evolution and nature of the human language faculty*, 42-64. Oxford: OUP.

Boeckx, C. (2015) *Elementary syntactic structures—prospects of a feature-free syntax*. Cambridge: Cambridge University Press.

Bolhuis, J.J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience* 11**,** 747-759.

Bruner, E., & Iriki, A. (2016). Extending mind, visuospatial integration, and the evolution of the parietal lobes in the human genus. *Quaternary International*, 405, 98-110.

Calabrese, A., and Woolley, S. M. (2015). Coding principles of the canonical cortical microcircuit in the avian brain. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3517– 3522. doi: 10.1073/pnas.1408545112

Cantalupo, C., and Hopkins, W. D. (2001). Asymmetric Broca's area in great apes. *Nature* 414(6863): 505–505. doi:10.1038/35107134.

Clarke, E., Reichard, U. H., and Zuberbühler, K. (2006). The syntax and meaning in wild gibbon songs. *PLoS ONE* 1: e73. doi:10.1371/journal.pone.0000073

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory* 2**,** 113-124.

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: The MIT Press.

Chomsky, N. (2005). Three factors in human language design*. Linguist. Inq.* 36, 1-22.

Chomsky, N. (2008). "On phases," in *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud,* eds. R. Freidin, C. Otero & M.-L. Zubizarreta. (Cambridge, MA: MIT Press), 133-166.

Chomsky, N. (2010). "Some simple evo-devo thesis: How true might them be for language?," in *The evolution of human language: Biolinguistic perspectives*, eds. R. K. Larson, V. Deprez, and H. Yamakido (Cambridge: Cambridge University Press), 45-62.The

Chomsky, N. (2012). The poverty of stimulus: unfinished business. *Studies in Chinese Linguistics* 33-1, 3-16.

Chomsky, N. (2013). Problems of Projection. *Lingua* 130-1, 33-49.

Chomsky, N. (2014). Language architecture and its import for evolution. Ms. MIT.

Darwin, C. (1859) On the Origin of Species, London: John Murray.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex.* London: John Murray.

Deacon, T. W. 1997. *The Symbolic Species.* New York: W. W. Norton & Company, Inc.

Di Sciullo, A-M. (2011). "A biolinguistic approach to variation," in *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of Human Language Faculty*, eds. A-M. Di Sciullo and C. Boeckx (Oxford: Oxford University Press), 305-326.

Di Sciullo, A-M. (2013). Exocentric Compounds, Language and Proto-Language. *Language and Information Society* 20, 1-26.

Di Sciullo, A-M. (2014). "Minimalism and I-Morphology," *Minimalism and beyond: Radicalizing the interfaces*, eds. P. Kosta, L. Schurcks and T. Radev-Bork (Amsterdan: John Benjamins), 267-286.

Dessalles, J.L. (2007). *Why we talk: The evolutionary origins oflLanguage*. Oxford: Oxford University Press.

Farrar, F. W. (1870) Philology & Darwinism. *Nature*, 1: 527-529.

Fernald, R. (2006). Casting a genetic light on the evolution of eyes. *Science* 313, 1914–1918. doi: 10.1126/science.1127889.

Fitch, TW (2010) *The evolution of language.* Cambridge: Cambridge University Press.

Gannon, P. J., Holloway, R. L., Broadfield, D. C., and Braun, A. R. (1998). Asymmetry of Chimpanzee Planum Temporale: Humanlike Pattern of Wernicke's Brain Language Area Homolog. *Science* 279(5348): 220–222. doi:10.1126/science.279.5348.220.

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207. doi: 10.1038/nature04675

Gould, S. J. (1976). "In defense of the analog: a commentary to N. Hotton," in *Evolution, Brain, & Behavior: Persistent Problems*, eds R. Masterson, W. Hodos, and H. Jerison (New York, NY: Wiley), 175–179.

Haimoff, E.H. (1984). "Acoustic and organizational features of gibbon songs," in *The Lesser Apes. Evolutionary and Behavioural Biology,* eds. H. Preuschoft, D.J. Chivers, W.Y. Brockelman & N. Creel. (Edinburgh: Edinburgh University Press), 333-353.

Harris, K. D. (2015). Cortical computation in mammals and birds. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3184–3185.

Hauser, M.D., Chomsky, N., and Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298**,** 1569-1579.

Hewes GW (1973) Primate communication and the gestural origins of language. *Current Anthropology* 14: 5.

Hornstein, N. (2009). *A theory of syntax*. Cambridge: Cambridge University Press.

Hurford, J. R. (2012). *The Origins of Grammar: Language in the Light of Evolution II.* Oxford: Oxford University Press.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends Cogn. Sci*. 3, 272–279. doi: 10.1016/S1364-6613(99)01333-9

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning and Evolution*. Oxford: Oxford University Press.

Jespersen, O. (1922). *Language*: *Its Nature, Development, and Origin.* London: George Allen and Unwin Ltd.

Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahasi M et al. (2013) Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*. Jun 6;498(7452):104-108.

Liu, WL, Gardner TJ, Nottebohm F. (2004) Juvenile zebra finches can use multiple strategies to learn the song. *PNAS* 101 (52): 18177-82. doi:10.1073/pnas.0408065101.

Marler, P. (2000). "Origins of music and speech: insights from animals," in *The Origins of Music*, eds N. L. Wallin, B. Merker and S. Brown (Cambridge, MA: MIT Press), 31–48.

Marshall, J.T., and Marshall, E.R. (1976). Gibbons and their territorial songs. *Science* 193**,** 235-237.

Mithen, S. (2014) The Cognition of Homo neanderthalensis and H. sapiens: Does the use of pigment necessarily imply symbolic thought? In: Akazawa, T., Ogihara, N., Tanabe, H. C. and Terashima, H. (eds.) The Dynamics of Learning in Neanderthals and Modern Humans. (Springer, London/New York), 7-15.

Miyagawa, S. (2010). *Why Agree? Why Move?: Unifying Agreement-Based and Discourse-Configurational Languages.* Cambridge, MA: The MIT Press.

Miyagawa, S., Berwick, R.C., and Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology* 4**,** Article 71.
Miyagawa, S., Ojima, S., Berwick, R. C., and Okanoya, K. (2014) The integration hypothesis of human language evolution and the nature of contemporary languages. *Front. Psychol.* 5. doi: 10.3389/fpsyg.2014.00564
Müller, Friederich Max. (1861) The theoretical stage, and the origin of language. In *Lectures on the Science of Language.* London: Longman, Green, Longman, and Roberts.
Newmeyer, F. J. (1991). Functional explanation in linguistics and the origins of language. *Language and Communication* 11, 3-28.
Newmeyer, F. J. (1998). "On the supposed 'counterfunctionality' of Universal Grammar: Some evolutionary implications," in *Approaches to the evolution of language*, eds. J. R Hurford, M. Studdert-Kennedy, and C. Knight (Cambridge, UK: Cambridge University Press), 305-319.
Nóbrega, V. and Miyagawa, S. (2015) The precedence of syntax in the rapid emergence of human language in evolution as defined by the integration hypothesis. *Front. Psychol.* 18 March 2015
Okanoya, K. (2002). Sexual display as a syntactical vehicle: the evolution of syntax in birdsong and human language through sexual selection. In A. Wray (Ed.), The transition to language (Oxford: Oxford University Press), 46-63.
O'Grady. W. (2005) *How children learn language.* Cambridge: Cambridge University Press.
Ogura, A., Ikeo, K., and Gojobori, T. (2004). Comparative analysis of gene expression for convergent evolution of camera eye between octopus and human. *Genome Res.* 14, 1555–1561. doi: 10.1101/gr.2268104.
Okanoya, K. (2004). Song syntax in Bengalese finches: proximate and ultimate analyses. *Advances in the Study of Behaviour*, 34, 297-346.
Ott, D. (2009) The evolution of i-language: lexicalization as the key evolutionary novelty. *Biolinguistics* 3(2), 255-269.
Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346:1256846. doi: 10.1126/science.1256846
Pinker, S. (1994). *The Language Instinct*. New York: Harper Collins.
Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences* 13, 707-784.
Progovac, L. (2015) *Evolutionary syntax*. Oxford: Oxford University Press.
Progovac, L. (2006). Fossilized imperative in compounds and other expressions. *Online Proceedings of the Inaugural Meeting of the Slavic Linguistics Society.* [http://www.indiana.edu/~sls2006/page6/page6.html.]
Progovac, L. (2008). "What use is half a clause?," in *Evolution of Langauge: Proceedings of the 7th International EVOLANG Conference*, eds. A. D. M. Smith, K. Smith, and R. Ferrer i Cancho (New Jersey: World Scientific), 259-266.
Progovac, L. (2009). "Layering of grammar: Vestiges of evolutionary development of syntax in present-day languages," in *Language Complexity as an Evolving Variable*, eds. G. Sampsom, D. Gil, and P. Trudgill (Oxford: Oxford University Press), 203-212.

Progovac, L. (2010). Syntax: Its evolution and its representation in the brain. *Biolinguistics*, 4.2-3, 1-22.
Progovac, L. (2012). Compounds and commands in the evolution of human language. *Theoria et Historia Scientarum* 9, 49-70.
Progovac, L. and Locke, J. (2009). The urge to Merge: Insult and the evolution of syntax. *Biolinguistics* 3.2-3, 337-354.
Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (ed.), *Elements of grammar: A handbook of generative syntax*. Dordrecht: Kluwer.
Samuels, C. (2015) Can a bird brain do phology? *Front*. *Psychol*. 6:1082. doi: 10.3389/fpsyg.2015.01082
Seed, A., and Tomasello, M. (2010). Primate cognition. *Topics Cogn. Sci.* 2, 407–419.
Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210, 801-803. doi: 10.1126/science.7433999
Spelke, E. (2003) What makes us smart? Core knowledge & natural language. *Language in mind: advances in the study of language & thought*, eds., Dedre Gentner, D and Goldin–Meadow, S. Cambridge, MA: MIT Press, 277–311.
Tallerman, M. (2007). Did our ancestors speak a holistic protolanguage? *Lingua* 117, 579-604.
Tallerman, M. (2016). Against the emergent view of language evolution. *Proceedings of EvoLang XI.*
Tallerman, M. and Gibson, K. R. (2012). "Introduction: The evolution of language", in *The Oxford Handbook of Language Evolution*, eds. M. Tallerman and K. R. Gibson (Oxford: Oxford University Press), 1-35.
Tattersall, I. (2009). "Language and the origin of symbolic thought," in *Cognitive Archaeology and Human Evolution,* eds. S.A. De Beaune, F.L. Coolidge & T.G. Wynn. (New York: Cambridge University Press), 109-116.
Wynn, T. (1998) Did *Homo erectus* Speak?" *Cambridge Archaeological Journal*, vol 8, 78-81.
Zhang, N. (2007). Root merger in Chinese compounds. *Stud. Linguist*. 61, 170–184. doi: 10.1111/j.1467-9582.2007.00131.