Language in Nature: on the Evolutionary Roots of a Cultural Phenomenon

Willem Zuidema

Abstract What could an evolutionary explanation for language look like? Here I review relevant evidence from linguistics, comparative biology, evolutionary theory and the fossil record, which suggest vocal imitation and hierarchical compositionality as the essential and uniquely human biological foundations of language. I also outline a plausible scenario for how human language evolved, and propose that language preceded, and facilitated the development of, other cognitive domains such as reasoning, the ability to plan, and consciousness.

1 Introduction

What distinguishes Man from beast? For all of human history we have been wondering about that question, and over the centuries we have learned to dismiss some of the popular answers. Humans might walk upright more than any other ape, have less hair, be better at long distance running, use tools more readily, have more advanced reasoning skills, be more aware of the thoughts of others or behave more cooperatively. But all of these features, it has turned out, are differences of degree and not of kind. One answer, however, has survived all serious scrutiny: humans have language. In other animals we find elaborate communication systems, sometimes with one or two properties vaguely reminiscent of language, but always differing radically in many other properties.

Although it is difficult to list the defining properties of language, there simply is no other animal that comes close to having something like human language, and, inversely, there is no human population that does not have it. Moreover, we use language typically for many hours per day, and language is involved in all parts of human life: in gossiping, shopping, education, politics, fighting, courtship, and

Willem Zuidema
Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherlandse-mail: zuidema@uva.nl

everything else. And, although much remains ill-understood, many scientists suspect that language somehow facilitates other cognitive skills that are dear to us: music, reasoning, consciousness, planning, mathematics and more. Hence, it is no overstatement to say that, from an evolutionary point of view, language is the most striking aspect of the human phenotype and cries out for an evolutionary explanation.

What could an evolutionary explanation for language look like? Libraries are filled with books on this issue, but many of the proposals are very speculative and, in fact, inconsistent with available evidence. It's worthwhile, therefore, to step back a bit and first consider some of the sources of information that could constrain the scenarios we might want to propose. The relevant evidence for evaluating evolutionary scenarios — consisting of particular starting and end points, and a mechanism that drives the steps in between — comes from many different fields. The end point, in our case, is the human capacity for language, and the obvious field to provide data is linguistics (although this field can offer less clear answers than we would perhaps wish). The starting point is the set of abilities of the last common ancestor that humans share with chimpanzees, our closest relatives. Our best guesses on these abilities come from a comparison of the abilities of other living great apes, i.e., from behavioural biology. The steps in between are largely unknown, but we find some hints in the fossil record. The mechanisms driving the evolution of language are also largely unknown, but evolutionary theory offers at least some constraints on the form of evolutionary scenarios. Finally, evidence on the abilities of more distantly related animals, such as songbirds, helps assessing the plausibility of these scenarios (by reasoning about 'convergent evolution' as explained below).

In this chapter I will survey some of these sources of information to get an idea what form an evolutionary explanation for the human-specific, and possibly language-specific, linguistic abilities should take. But before we embark on a discussion of the anatomy and abilities (section 4 and 5) of humans and other animals, we must first consider how we can apply the standard approach from evolutionary biology — the comparative method — to a culturally evolved system like language (section 2) and why we don't take one of the elaborate theories from linguistics as our starting point (section 3).

2 The comparative method in the light of cultural evolution

In investigating the evolution of language we will of course pay special attention to those traits that are unique to humans among the apes — the human-specific traits — which are likely to have evolved since that common ancestor. Moreover, we might want to distinguish, as well as we can, between traits that emerged in human evolution independently from their function in language and those that are in fact language-specific. However, it would be a mistake, for three reasons, to limit our attention to such uniquely human or uniquely linguistic abilities alone. First, one of the most successful approaches in biology for understanding the evolution of

particular traits is in fact based on trying to identify commonalities between different species: by comparing many different species and considering the evolutionary relationships and similarities and differences in their ecology, biologists can try to reconstruct the evolutionary history of a trait, and attribute commonalities between two species to homology (the two species inherited the trait from a common ancestor) or analogy (the trait evolved independently in both species due to similar selection pressures, a process known as 'convergent evolution'). Applying such a comparative method to language turns out to yield a more powerful approach than many armchair theorists stressing the uniqueness of language realized (Fitch, 2005).

Second, as explored in other chapters of this book, language is a rather unique system in nature, because it is transmitted culturally from generation to generation and can undergo cultural evolution. For research on the biological evolution of language abilities this is a very relevant fact, because it radically changes what counts as evidence for one theory or another. In particular, it is important to realize that not every difference between humans and other apes is equally interesting, not even if we limit ourselves to traits that are demonstrably relevant to language. To see why, consider that when we compare the vocalizations or learning abilities of any two species, we will necessarily find many differences that are accidental in some sense. In the case of language, we know that the cultural evolution process, where languages adapt to language learners, will result in languages that reflect such accidental properties. The very fact that the peculiarities of languages and those of humans 'match' is thus expected even in the absence of biological adaptation.

Adaptations are traits that evolved because they conferred a fitness advantage, that is, because individuals with the traits on average obtained more offspring than individuals without them ('fitness' of an individual in evolutionary biology is defined as the expected number of offspring of that individual). When looking for biological adaptations for systems like language that can undergo cultural evolution, we need to look for differences in traits that still have effects on fitness after the process of cultural evolution has unrolled. It would be a mistake to classify as an adaptation every uniquely human trait that is more useful for learning and using human language than an ancestral trait, because the ancestral trait might in fact have been equally good for learning an ancestral language and the good match between humans and modern language a result of cultural rather than biological evolution. Unfortunately, many discussions of language in a comparative perspective make that mistake. For example, Pinker & Jackendoff (2005) list many properties of speech perception that they take to be unique to humans and adaptations, including differences in preferred category boundaries for humans and nonhuman animals and the fact that human neonates have a preference for speech sounds. These features might be unique for humans, however they are more likely accidental features that language adapted to than biological adaptations for language.

Third, another consequence of the cultural evolution of language is that there is no one-to-one correspondence between the 'human capacity for language' and the features of individual languages. Human children can learn any natural language, but languages can be very different and not all features of the human language capacity are necessarily exploited by any particular language. Similarly, any particular

communication system found among a population of non-human apes might not reflect their capacities to the full. As human languages have evolved culturally to adapt to features of the human brain, the possibility remains that human languages reveal previously hidden talents of the ape brain: features shared with other apes even if they have left no observable effects on ape communication.

3 Linguistics and language evolution

Investigations of the evolution of language naturally start with the question: what is language? The good news is that, at a very general level, linguists all agree: languages are complex, acquired systems of conventions about relations between forms (e.g., spoken or signed utterances) and meanings. The forms are built up by combining elementary units from a basic inventory (phonemes, syllables, hand shapes), and utterances are built up by combining meaningful units (morphemes, words, gestures) into phrases, sentences and discourse, following rule-like patterns. Every human population has language, and in practice, linguists have no difficulty determining which behaviors in an unknown culture count as language, and which as nonlinguistic sounds (e.g., music) or gestures (e.g., dance).

However, the bad news is that the consensus ends at this very general level. The moment we want to make more precise what language in modern humans exactly is, controversies pop up everwhere. For instance, what are those elementary units of form? Even when describing a single language, like English, disagreements abound. Some theories assume the elementary units are phonemes, others that the atomic level is that of 'distinctinctive features' (e.g., Chomsky & Halle, 1968). More recently, a popular position is to take larger units — syllables or exemplars — as atomic (Levelt & Wheeldon, 1994). And this is only the beginning; much more controversy surrounds more complex units, further removed from direct observation, such as morphemes or grammar rules.

The lack of consensus is even more apparent when considering the full diversity of languages in the world. Languages differ beyond imagination (Evans & Levinson, 2009). Some languages build up incredibly long words that convey the meaning of a complete sentence in English; some languages have an almost completely free word order, but mark with a complex system of inflections the roles that various words in a sentence play. Other languages obey strict word order rules, but lack any kind of word morphology, including even plural markers like -s in English. Some language use only a handful of phonemes, others have well over a hundred distinctive atomic sounds. The usefulness of even the most basic concepts of linguistics — 'word', 'phoneme', 'subject', 'rule', 'category' — is regularly questioned in the description of one language or another.

Nevertheless, comparison with other animals does quickly make clear that human language is qualitatively very different from any other communication system in nature, even if a convincing, integrated theory of how language works remains elusive. There are interesting questions to be asked about why linguistics is in this

state, and why descriptive and theoretical linguistics seem to have so little to offer to solving questions about the evolution of language. I suspect that cultural evolution, and the fact that languages have adapted to the messy idiosyncracies of the human brain, has much to do with it. For the purposes of this chapter, however, the best way forward is to take a pragmatic approach and focus on those aspects of language and speech where empirical research comparing humans and other animals has revealed important qualitative differences — these differences are candidates for the adaptations for language (and speech) that we are after.

4 Anatomy and language

4.1 Speech Production

When we look at the anatomy of the human speech production and perception apparatus, we see a strong continuity with the other great apes and even the broad class of mammals. To produce sounds, many mammals, like humans, let air flow from the lungs through the larynx, the throat and the nose and mouth. The larynx contains special membranes, the vocal folds, which vibrate in the air flow and can be tightened or loosened to produce higher or lower pitched sounds. The cavities between larynx and the lips together form the vocal tract, which effectively filters the mesh of sounds created by the larynx, by reinforcing some frequencies (resonances) and attenuating others. Three features of the human anatomy used in speech production stand out (see Figure 1): the fact that the larynx is very low in the throat, that humans, unlike all other apes, have no air sacs, and that humans have detailed and rapid control over the shape of the vocal tract (see Crystal, 1997, for an accessible review of the human anatomy involved in speech production and perception).

The human larynx is high in the throat in babies (allowing them to breathe and drink at the same time), but descends to the lower position as they get older. In males, there is a second descent of the larynx during puberty. The position of the larynx is very relevant for speech as it determines the length of the vocal tract, and the size and shape of the vocal tract in turn determine the quality of the sound that comes out. Naturally, human vocal tracts are ideal for producing human speech sounds, but is the system as a whole 'better' in some way? Philip Lieberman (1984) has argued that the descended larynx allowed a much richer repertoire of speech sounds, and could thus confer a fitness advantage that offsets the disadvantage of an increased probability of choking. Lieberman went as far as claiming that this was the crucial innovation in the evolution of language. Although recent modelling work (de Boer, 2010) upholds his claim that the human vocal tract is optimal for producing a range of distinctive sounds, the effects are small and unlikely to be the crucial factor in the evolution of speech.

Moreover, the permanently descended larynx turns out to be not uniquely human but is also found in red deer and other species without language (Reby et al.,

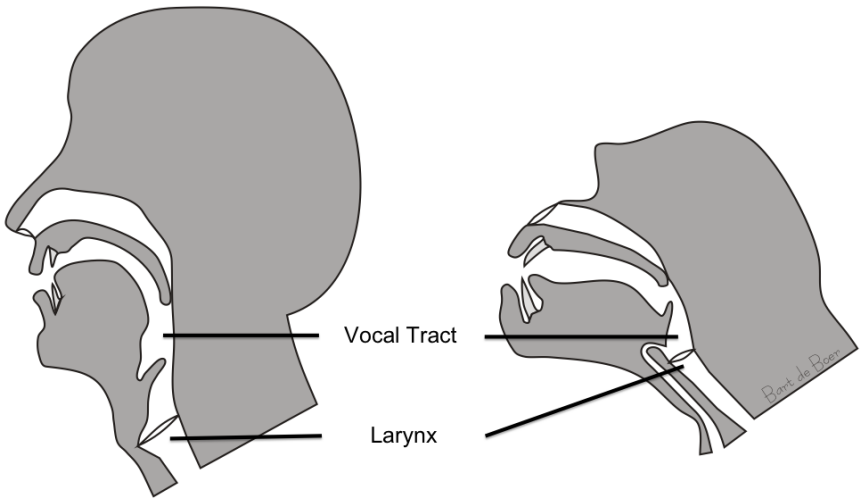

Fig. 1 Human (left) and chimpanzee (right) vocal anatomy differ in three important respects: in humans, the larynx is positioned lower in the throat and the tongue is rounder, yielding a vocal tract with equivalent controllable back and front cavities; humans have no air sacs attached to the larynx (chimpanzee air sacs are attached to the larynx through the narrow tube that can be observed in front of the larynx), further improving the range of sounds that can be produced; humans have voluntary control over the movements of the vocal folds. Diagram courtesy of Bart de Boer; based on FMRI data in Fitch (2000).

2005), strongly suggesting that there is at least one other biological function for a descended larynx. Fitch (2000) suggests this other function might be size exaggeration: with a larynx low in the throat one can make sounds that otherwise only much large animals could make. Finally, there is at least one mammal without a permanently descended larynx that is, under exceptional circumstances, very good at imitating human speech: recordings from the harbor seal Hoover, raised in a fisher man's bath tub, contain a few intelligible sentences (Ralls et al., 1985). Hence, the position of the larynx might very well have been a target of natural selection for speech once rich languages had emerged among hominids, but it is unlikely to be a crucial factor in the emergence of a rich language in the first place.

Much less attention has been given to the absence of air sacs. All other apes have such sacs: cavities attached to the larynx that can range from modest in size (chimpanzees) to clearly visible balloons in the neck (gorillas). It's clear that air sacs have an acoustic effect on the vocalization produced, and various researchers in the last century have formulated the hypothesis that humans lost air sacs because of a detrimental effect on speech comprehensibility. In recent modelling and experimental work, de Boer (2009) confirms the suspicion that air sacs have such a detrimental effect. However, as with the descended larynx, the effects are not enormous. The

loss of air sacs is likely to have been affected by evolutionary pressures for speech, but it is unlikely to be the key event that set all the rest in motion.

One fascinating aspect of air sacs is that they have left traces in the archeological record: the shape of the 'hyoid bone' in the throat correlates with the presence or absence of airsacs, and this bone is occasionally but very rarely fosselized in ape and hominin fossils. Based on the few findings reported, we can make a rough estimate of the disappearance of air sacs: Australopithecus afarensis, an human ancestor that lived about 3.3 million years ago still had air sacs (Alemseged et al., 2006), while Homo heidelbergensis, who lived some 600,000 years ago did not (Martínez et al., 2008; for a review of homin fossils and approximate timelines, see Jones et al., 1992).

A third anatomical oddity of human speech is the extremely rapid control over vocalizations, with precise, millisecond-level synchronisation of movements at distant places in the vocal tract, from the larynx to the lips. On high-speed x-ray films of the human vocal apparatus one can see complex, extremely fast and accurate movements that tongue, lips and other articulators make when producing a string of words. Although it is difficult to quantify, nothing comparable has been reported in the vocalizations of other primates. In song birds, however, we do see extremely fast and complex vocalizations as well, with precisely timed simultanous movements in syrinx (the bird equivalent of the larynx) and beak. An open question is whether birds, like humans, deliberately manipulate the resonance frequencies of the vocal tract (e.g., by moving the tongue or by opening air sacs), but preliminary evidence (Ohms et al., 2010) seems to point in this direction. In both humans and song birds, but not other primates as far as we know, forebrain regions seem to be involved in the control over vocalizations (Deacon, 2000). Combined with the fact that in other primates we find only limited vocal repertoires and relatively simple and slow vocalizations, the findings on the extremely versatile articulatory control in humans suggests that evolutionary innovations could have been essential for the high rate of information transfer through speech that modern humans are capable of.

In short, there are some likely biological adaptations to the anatomy of the vocal tract that would have improved communication through speech, but none, it seems, that would have been necessary for language to emerge in the first place. There are some further likely biological adaptations in vocal control; these, in contrast, might have been essential for language, in the vocal-auditory channel at least, to confer a fitness advantage.

4.2 Speech Perception

In human speech perception, the relevant anatomical structures seem even more similar to what is common among mammals. Although the shape of the outer ears vary widely, the middle and inner ears of other (land) mammals — exquisitely complex organs — are very comparable (although the shape of the outer and inner ears might be responsible for increased sensitivity of humans in the frequency range needed

for speech perception, Martínez et al., 2004). Behind the ear drum, we find the same hammer, anvil and stirrup that conduct the vibrations to the cochlea, where they are translated into neural activation patterns in the Organ of Corti. The auditory nerve then transports these signals to higher processing levels in the brain.

Behaviorally, however, human speech perception does seem special. In the 1960s, pioneering research by Alvin Liberman and colleagues (Liberman et al., 1967) revealed that subjects could still perceive differences between two different speech sounds at high levels of background noise where differences between other sounds are lost. On the other hand, subjects did not perceive much difference between two versions of the same phoneme, even if the physical difference was of similar magnitude. This phenomenon, that differences between categories are perceived much more clearly than differences within a category, is called "categorical perception". Liberman and colleagues further found that subjects especially perceive physically different sounds as similar if they are produced with similar movements. This led the authors to propose a "motor theory of speech perception" which states that speech is perceived in terms of the movements necessary to produce it.

Together, these results led to the "speech is special" hypothesis, which states that humans are biologically specialized for speech perception. Much interesting research has been published since, comparing human and non-human perception, of linguistic and non-linguistic stimuli. Where human speech perception was found to differ from non-speech or non-human perception, such findings were often claimed to be 'adaptations' (e.g., Pinker & Jackendoff, 2005). However, as we saw before, the close fit between language (in this case, speech sounds) and human abilities (in this case, speech perception) is not in itself conclusive evidence for biological adaptation.

One recent finding by Smith & Lewicki (2006) is telling in this respect. They considered the different ways in which neural firing patterns can encode auditory information; such neural encoding is ultimately what the inner ear does with auditory input, to send information for further processing to the brain. Some neural codes are very efficient for one type of acoustic input, others for other types of input. Smith & Lewicki discovered, to their surprise, that the code used in the inner ears of cats (as derived from neurophysiological studies) appears optimised for human speech. As it is very unlikely that cat's hearing has indeed been adapted to human speech since their (in evolutionary terms) recent domestication, the only sensible explanation is that the causality is the other way around: human speech exploits those sounds that the mammalian auditory system can most efficiently process. And, indeed, in the same study, Smith & Lewicki find that the same encoding is also optimal for the sample of sounds they created from a mixture of ambient sounds (water flowing, cracking twigs) and animal vocalizations.

Hence, although the empirical discoveries by Liberman et al. still stand in broad outline, Smith & Lewicki (2006) and many other studies since the 1960s have put the original interpretation into perspective. On the perception side, it looks like humans make use of a biological apparatus that hasn't fundamentally changed from our prelinguistic ancestors, although there are many human peculiarities that languages would have adapted to culturally.

5 Design Features of Language

5.1 Cultural transmission and Vocal learning

When we turn from aspects of human anatomy to less tangible, structural aspects of language, a list of 'design features' of human language by Hockett (1960) is a useful starting point. This list has since been a focus for research comparing language with natural vocalizations in other animals, in particular the four main design features: cultural transmission, symbolism, duality of patterning and hierarchical phrase-structure. I will first focus on cultural transmission, which refers to the fact that languages, within the constraints set by our biology, are conventional systems that persist through time by repeated learning. This is true for the elementary sounds of spoken languages, for the elementary shapes and movements in sign language, and for all the grammatical rules and constructions. Cultural transmission is not unique to language or humans — we also observe it in, e.g., music and bird song — but rare among primates and a key qualitative feature of language.

Focusing on sounds, cultural transmission is possible thanks to the ability for vocal imitation: the ability to relate perceived sounds back to the articulatory movements that can produce it. Vocal imitation is, as far as we know now, absent or very limited among other primates (Janik & Slater, 2000), with the possible exception of gibbons. Early language-training studies with apes famously failed to get the apes to produce any speech-like sound, and natural vocalizations in monkeys and apes appear to be innate (i.e., develop independently from exposure to those sounds, cf. Seyfarth & Cheney, 1997). Among mammals, the only groups other than humans with vocal imitation are seals, bats, elephants, dolphins and whales. For instance, humpback whales sing long songs, that are shared among members of one population of multiple generations, but differ from population to population and gradually change over time (Payne & McVay, 1971). Among birds, finally, there are very many vocal learning species, but they are limited to three groups: songbirds, humming birds and parrots.

Although it is not completely clear what the criteria for true vocal learning are (evidence for vocal learning in the mentioned species sometimes comes from experiments with controlled training stimuli, sometimes from field observations of imitation or cultural transmission; Janik & Slater, 2000), it does seem clear that advanced vocal learning is rare but found in multiple species scattered over the evolutionary tree of life. This presents a wonderful opportunity to investigate the possibility of convergent evolution: in various branches of the evolutionary tree similar solutions evolved for similar problems. Why is vocal learning rare? What are the difficulties or disadvantages preventing most species from having the ability, and what are the advantages that drove its evolution in the species, including humans, that do have it?

The question about difficulties is all the more pertinent, because vocal learning — from a computational point of view — is not something particularly complex. Some existing computer models of vocal learning might help to find an answer. For

instance, Westermann & Miranda (2004; see also Oudeyer, this volume) present an elegant model of neural structures that can learn mappings from articulatory movements to sounds, and vice versa, and thus implement vocal learning. The model consists of two neural maps, one representing motor activity and the other perceptual input. It assumes a babbling phase, where the learner initially produces sounds at random, and 'articulatory feedback', meaning that he can hear himself. Given those assumptions, the model shows how (Hebbian) connections between neurons in both maps can come to encode the relation between movements and sounds. When learning is complete, the model can be prompted with just a sound and then produce, in the motor map, the pattern of activity needed to generate that sound. Hence, it shows the potential for vocal imitation. A straightforward extension of the model with a visual map also makes correct predictions about the influence of seeing lip movements on the sounds perceived (the so-called McGurk effect).

Comparing the model to real brains, there are two clues to what might make vocal learning difficult in reality. First, the connections between maps in the model are bidirectional: the same connections are used for predicting sounds given motor parameters as for vocal imitation. In real brains, the ability to predict sounds given motor activity is likely to be common among animal species, but because neurons are not bidirectional, a dedicated pathway might be necessary to also learn the inverse mapping. Second, motor and sound map activity are static in the model; in real brains, the motor maps will already be involved in planning for the next vocalization by the time the articulatory feedback arrives (Dave & Margoliash, 2000). This thus necessitates a memory-motor map in addition to the motor and perceptual map in the model, and a dedicated pathway to transfer information from the memory-motor map back to the actual motor map for the production of vocal imitations.

Brain research on song birds has indeed found evidence for such dedicated pathways; intriguingly, the solutions found in the independent evolution of vocal learning in song birds, humming birds, parrots and humans appear to be very similar (Jarvis, 2004). Jarvis (2006) argues that this is indeed a case of convergent evolution, and observes that there is one thing the species with vocal learning have in common that distinguishes them from many non-vocal learning sister species: they are at the top of the food chain and often have few, if any, predators. This observations needs more systematic research, but it could be a key factor, because vocal learning of a complex repertoire of sounds requires practice. Practice sessions, where infants spend time and energy (that could otherwise be used for more direct ways to increase survival) and make noise that attracts predators, might simply not be a viable option for species that are under predation pressure.

Given that vocal imitation is possible, but not trivial and likely to be costly, the question arises what the evolutionary advantages could be in song birds and, ultimately, in humans. Jarvis surveys a number of popular ideas in the literature, including the idea that learned vocalizations allow for individual identification and for cultural adaptation to diverse habitats where different types of sounds might transmit better. Coming from bird song research, he favors the hypothesis that the variability allowed by vocal learning played in a key role in mate attraction; given

the attractiveness of human singers to members of the opposite sex this is indeed a serious candidate selection pressure in human evolution as well.

However, the increase in the number of distinct sounds that can be produced could also have played a different role in humans than in other vocal learners, in particular in 'semantic communication'. Jarvis is skeptical of a role for semantic communication in the evolution of vocal learning, and correctly points out that in song birds signals that carry meaning (like food or alarm calls) tend, in fact, not to involve vocal learning, while vocally learned song has no referential role. However, modern human language is very different in this respect, and Jarvis might stretch the songbird analogy a little too far here. In any case, I see no reason why the selection regime that allowed large, meaningful vocabularies to emerge in humans could not have played a role in the emergence of vocal learning in the first place.

In conclusion, vocal learning is rare trait in nature but crucial for the spoken language. The comparative record provides some clues about the questions as to why we have this trait and many other animals do not. Our position as a top predator might have removed selection pressures against it, while the need for a great variety of sounds in communication — useful in mate attraction or semantic communication — might have provided selection pressures for it. Firm conclusions, however, cannot be drawn at the current state of knowledge and alternative (but not necessarily contradictory) theories exists. For example, Lachlan & Slater (1999) propose, based on a mathematical model, a "cultural trap hypothesis" which states that once vocal learning has emerged and a variable repertoire is used in a species, for whatever reason, vocal learning is favored over innate vocalizations. Oliphant (1999) proposes that the difficulty of identifying the intended referents in learning a lexicon was a crucial obstacle in the evolution of cultural transmitted semantic communication in other species.

Perhaps genetic evidence, as is now starting to emerge (see Dediu, this volume) will play a role in the future in understanding the evolution of vocal learning. FoxP2 is a gene involved in speech and language, as discovered in the study of a family with a heritable disorder affecting several speech, language and motoric abilities. Through the careful work of Vargha-Kadem and colleagues (Vargha-Khadem et al., 1995) it has become clear that the gene is not specific for language (as proponents of an extremely modular view of the mind were perhaps hoping), although it does indeed seem to affect linguistic abilities over and above the indirect effects one can explain from effects on general intelligence and motoric abilities. Interestingly, the same gene also plays a role in vocal learning in birds (Haesler et al., 2004). Studies of variants of the gene in other species, including the extinct Neandertals (Krause et al., 2007), are starting to provide a fascinating look on the evolution of the gene, but given the many unknowns about the exact function of the gene it is too early to directly relate it to scenarios of the evolution of vocal learning and language.

5.2 Symbolism and Arbitrariness

A second feature of natural language that is often said to be unique is its 'symbolism', but this term can mean various things. One aspect of symbolism, featuring in most definitions, is that the relationship between the words or morphemes in a language and what they refer to is arbitrary. Thus, there is nothing in the sound of words like 'sleep', 'green' or 'democracy' that is in any way similar to what these words denote. Even onomatopeia — words that do mimick the sound they describe, such as 'cock-a-doodle-do' in English — are to a large part conventionalized, as can be seen from the fact that the same rooster's calls are referred to as 'kukeleku' or 'cocorico' in other languages (Dutch, Italian). Thus, there is no doubt that humans have the ability to assign arbitrary meanings to arbitrary sounds, and they do so all the time: adult native language users typically know many tens of thousands of words (Bloom, 2000). In that massive vocabulary, some words sound or look somewhat like what they denote, but the vast majority of word-meaning mappings are arbitrary (e.g., Tamariz, 2005).

This 'arbitrariness of the sign' is a feature of natural language that is cherished by many linguists, but arbitrariness per se might be less relevant from the comparative and evolutionary point of view than has often been assumed. There are many alarm call systems — in birds, primates, rodents — where particular sounds denote particular predators (or better, perhaps: denote the appropriate response to the presence of a given predator) and where there seems to be no relation between the sound and its meaning. Learning these associations is common too: although the production of calls is typically thought to be innate, the interpretation of calls is somewhat flexible, and different species of monkeys are known to be able to learn to interpret each others alarm calls (e.g., Zuberbühler, 2002). Moreover, arbitrariness is also not the all-or-nothing phenomenon that it has often been taken to be. In human sign languages (known to often be much more iconic than spoken languages; Frishberg, 1975) and ape gesturing (e.g. Tomasello et al., 1997), it has often been observed that gestures that start out as iconic, can gradually become more and more arbitrary (i.e., for an external observer the original iconic relationship between gesture and meaning is less obvious or even unobservable in later stages).

Other features of human vocabularies might be truly unique to humans, but also don't necessarily point to language-specific adaptations. Quantitatively, the readiness with which humans acquire a vocabulary is remarkable. Children start understanding and using their first words around their first birthday, and after a slow start, are estimated to learn 10 words a day between age 2 and 6, reaching a vocabulary of about 14,000 words by age 6, which further increases to perhaps as much as 60,000 words at high school age (O'Grady, 2005). These numbers dwarf any vocabulary size found in non-human animals, where chimpanzee gesture repertoires are estimated to contain at most a few hundred signs. The record-holder in vocabulary size, as established in a controlled experiment, is held by a dog: border collie Rico can recognize about 200 names for objects (Kaminski et al., 2004). Thus, there is a huge quantitative gap between humans and other animals, but this does not prove that humans have language-specific mechanisms for word learning and usage. The gap

could also be an indirect consequence of differences in reasoning abilities, in particular in abilities to reason about the intentions of others (Oliphant, 1999; Bloom, 2000). Moreover, it could be an indirect effect of evolutionary innovations in grammar (as discussed in sections 5.3 and 5.4): most words are learned concurrently with the grammar of a language, and the grammatical context provides additional clues to word meaning (Cruse, 1986).

Returning to the qualitative differences, another well-known observation about word learning is a "mutual exclusivity bias" in children: a preference for 1-to-1 mappings from words to objects, without synonymy (several words with the same meaning) and homonymy (identical words with multiple meanings). Interestingly, the mentioned border collie Rico seems to share this bias with humans: when confronted with a novel word, he was more likely to associate that with an object that he did not already know a word for. This suggests we do not need to assume a language-specific adaptation to explain this bias in humans, but can rely on general cognitive and communicative processes.

Finally, in most definitions symbolism is more than just the arbitrariness of word meanings. Harnad (2003) defines a symbol as an object that not only has an arbitrary meaning, but is also part of a symbol system. A symbol system, in turn, is a system of symbols and rules, where rules apply to symbols regardless of their meaning. Thus, the word 'cat' is a symbol, not just because its sound is in no way similar to the animal it denotes, but also because it participates in a system of rules and many other symbols: the English language. The rules of English (such as those that put the determiner 'the' in front of it, or the plural marker '-s' behind it) apply to it because of its syntactic category ('noun') and not because of what it means. The part-of-a-system requirement in Harnad's definition means that it makes no sense to speak of symbols in isolation (as Harnad himself point out); therefore, it seems to me that symbolism is an inseparable consequence of compositionality, which I will discuss in the next section.

Using similar arguments as Harnad, Paul Bloom (2004) has warned against overinterpreting the analogies between human word learning and the 'words' learned by dogs and other animals. He points out that the research with animals has not demonstrated that they can combine words for objects with all kinds of action words. This, however, is again the part-of-a-system requirement, and, I would argue, inseparable from compositionality.

For many scholars, there is even more to the symbolic nature of language than the part-of-a-system aspect. They feel there is something special about the relation between human words and the concepts they symbolize, but typically do not give a precise definition of that special, symbolic relationship. This doesn't stop many of them from speculating about the relevance of the appearance of art, from about 100,000 years ago, in the archeological record for scenarios of language evolution. This, they argue, is a strong clue for language, as art, like language, requires symbolic thought. I am skeptical about the confidence with which they argue for this point: it is hard to imagine any hominin species with only part of the human suite of cognitive abilities, but that is exactly our job. I don't see a priori reasons why art without language would be harder than language without art, and the comparative

record has unfortunately little to say on this issue (see Botha, 2008, for a detailed critique of several such proposals).

5.3 Duality of patterning

To a first approximation, sentences in languages are built up from meaningful words (or rather: morphemes), and words are built up from meaningless phonemes. Although the situation is more complicated than that, it seems fair to say that human languages employ at least two combinatorial systems: a combinatorial, phonological system that regulates which basic sounds can be combined into possible words, and a compositional, semantic system that regulates how words and their associated meanings can be combined to give sentences and compound meanings. Both combinatorial systems generalize to unseen sequences: we can interpret sentences we have never seen before, and distinguish impossible from non-existing but possible words in our native language (e.g., the French word "pluie" is an impossible word in English, because of the onset /plj/, while the Dutch word "vonk" is a nonexisting but perfectly possible word in English). Hockett (1960) used the term 'duality of patterning' for the marriage of a combinatorial phonology and a compositional semantics (somewhat earlier, Martinet, 1949, had already made the same observation using the term 'double articulation'.).

Most animal vocalizations, in contrast, are holistic: a single vocalization has a particular function, but there is no sense in which we want to analyze the vocalization as built-up from components that are reused in other vocalizations. It is interesting to look at the exceptions to this general statement, where there are three aspects to pay attention to: are vocalizations built up from several elementary units? Do these vocalizations have a referential meaning? And is the meaning of a combination somehow a function of the meaning of the parts? As always in biology, we find an enormous variety in nature and we do observe that combinatorial phonology and compositional semantics have their echos in other animals' communication systems. However, there are important qualitative differences and the presence of both in one species has, as far as I know, only been attested in humans.

When we only focus on the combination of vocal elements, there are in fact still quite a lot of examples in primates and cetaceans, and especially among songbirds. In many song birds we find distinct repertoires of basic elements that can be combined in various ways but according to quite strict rules. A good example is chaffinch song: Riebel & Slater (2003) describe the repertoire of a population of chaffinches, and the rules that govern the structure of the songs. Each (male) bird sings two or three different songs, and each song follows a stereotypical AxByF or AxByCzF pattern. Elements in the A, B and F part are repeated a varying number of times, but the elements that make the x and y part, called transitional elements, are never repeated although they can be omitted. Extremely similar A elements can be found combined with different Bs and Cs. From such findings, it has become clear that chaffinches have a combinatorial system in place. It differs, however, qualitatively

from what we find in language. Most importantly, the songs do not convey a referential meaning: songs have a function in attracting females and defending the male's territory against rivals, but the message to them is always the same: come here! or go away!. Variations in, for instance, the number of repetitions do not change that message. That means we cannot speak of duality of patterning, as there is only one system of combination, and we can perhaps not even speak of combinatorial phonology, as the term 'phonology' is usually reserved for meaning-carrying vocalizations/gestures. It also potentially limits the usefulness of looking at bird song for understanding the evolution of combinatorial phonology, because a crucial constraint that presumably operated in the evolution of phonology — the system must remain useful for encoding and decoding information — was missing in bird song evolution.

Combination of vocalizations that do carry meaning is, in contrast, very rare. Arnold & Zuberbühler (2006) describe a communication system in putty-nosed monkeys that fits the bill. The monkeys use distinct, loud alarm calls to warn each other of predators: they emit the so-called pyow call when a leopard is detected in the vicinity, and a hack call for eagles. Additionally, the monkeys sometimes produced pyow-hack sequences, consisting of 1-3 pyows followed by 1-4 hacks. These sequences are produced in response to both eagles and leopards, and are typically followed by the whole group of monkeys moving to a different area of the forest. In their study, Arnold & Zuberbühlerdemonstrated experimentally that the pyow-hack sequences indeed mean something different than the individual pyows or hacks. They played leopard growls to 17 groups of monkeys, each consisting of a single adult male, with several females and their offspring; in about half of the groups, the adult male responded with a pyow-hack sequence, and those groups were found to have moved significantly further away 20 minutes later than the groups that only responded with pyows — the leopard alarm call. Putty-nosed monkeys thus have a rudimentary form of combinatorial phonology: elementary sounds, used to denote the two predators, are reused to form a third signal which roughly means: "let's go". But the putty-nosed monkeys do not exhibit compositional semantics, as the meaning of the combined signal is not somehow derived from the meanings of its component parts.

In another study, however, Zuberbühler (2002) did find rudimentary compositionality: Campbell monkeys also have a system of alarm calls for various predators, and aside from the usual unitary calls, also sometimes produce sequences of two calls. The first call is a so-called 'boom', and modifies (weakens) the meaning of the second call, a leopard or eagle alarm. Zuberbühler experimentally demonstrated that Diana monkeys — another monkey species living in the same habitat and eavesdropping on the Campbell calls — withhold the usual response to the alarm calls if they are preceded by the boom. The meaning of the whole, it seems, is thus a function of the meanings of the parts. That places the Campbell monkeys in an odd class of species for which some form of compositional semantics has been attested. The only other members of that class are humans and some species of bees, who convey the location of a food source through a dance where two components of the form

(direction, length) map onto two components of the meaning (direction with respect to the sun, distance from the beehive; von Frisch, 1974).

The results from Zuberbühler and colleagues are important, but they do only demonstrate very rudimentary forms of either side of the duality of patterning. They add to the evidence that combining two signals to mean something new, and combining meanings to create a compound meaning, are feats that do not necessarily require a language-adapted brain. Although it's novel to see such evidence in natural communication systems, we already knew from trained apes, dogs and other domesticated animals that a rough combination of the meaning of two sounds is within reach of those animals (e.g., Truswell, in prep.). Compositional semantics in natural language, however, is quite a bit more sophisticated. For instance, in many languages word order is a crucial variable, such that 'dog bites man' means something different from the sentence in the reverse. Such a phenomenon has never been attested in any non-human animal.

We shouldn't be surprised if a monkey communication system is soon discovered that combines the tricks of the Campbell and putty-nosed monkeys, and thus provides a rudimentary duality of patterning. But exciting as such a finding would be, such a rudimentary form would not tell us much about the evolution of duality in human language. What we would really like to know is whether the ability for an extensive duality of patterning is already lurking inside the primate brain, but species other than humans lack the motivation to use it, or whether we need dedicated brain structures to be able to process it. To answer these questions, we need to know much more about the neural and cognitive mechanisms that underlie duality of patterning in humans. From a purely computational point of view, it is hard to see why compositional semantics would be particularily difficult for a monkey, ape or bird brain that can already readily process combinatorial conceptual structure (such as needed for planning, vision and social cognition) and combinatorial signals. The only obvious difficulty derives from the fact that compositional semantics requires combinatory operations to apply to representations of meaning and representations of form in synchrony, in such a way that the system become bidirectional: language users must be able to compute the meanings of a given form, or the appropriate form for an intended meaning using the same rules of language. There might be difficulties, though, in integrating pieces of information that are processed in different parts of the brain, similar to what we saw in the case of vocal learning. More research is necessary on the particularities of how primate brains handle conceptual structure and communicative signals.

Meanwhile, we can ask what the evolutionary costs and benefits of duality of patterning could be, assuming that brains can implement it. Martin Nowak and others studied a number of interesting mathematical models that bear on this question (e.g., Nowak & Krakauer, 1999; Zuidema & de Boer, 2009). The basic insight underlying this work is simple: combinatorial systems can convey many more messages than holistic ones with the same number of elementary units; e.g., a system with 10 nouns and 10 verbs can handle 10x10=100 distinct noun-verb combinations, whereas an holistic system of the same size can only convey 10+10=20 distinct messages. How many messages do we need to convey? Nowak & Krakauer reason

that species will differ in how many distinct signals they will want to communicate with; we will call that number N. Now, it is reasonable to assume that the number of distinct elements that can be learned, remembered or distinguished from each other is limited to some number M. Moreover, we must assume a cost to using a combinatorial strategy (for instance, because learning 20 nouns and verbs is more difficult than learning 20 holistic signals, or because combinatorial strategies use up more memory and energy), but for a given M and cost there is always some number N at which combinatorial strategies will outperform holistic strategies. Nowak and colleagues therefore suggest that a possible explanation for why humans have combinatorial phonology and compositional semantics is that they were more cooperative and wanted to communicate more distinct messages than other primate species. In other words, the human N is above the threshold for combinatoriality, while the chimpanzee N is not.

A more difficult question is how natural selection could have driven the transition from holistic systems, to communication codes with either combinatorial phonology or compositional semantics, or both. In Nowak & Krakauer (1999) and related papers, Nowak et al. show that in species that have both a holistic and combinatorial system in parallel, the evolutionary dynamics will, under reasonable assumptions, always lead to using the combinatorial system more and more. However, Zuidema (2003) and Zuidema & de Boer (2009) argue that assuming two systems in parallel makes for a rather unrealistic scenario, and show that in a single system optimization for noise robustness can yield systems with both rudimentary compositional semantics and combinatorial phonology.

In sum, the extensive duality of patterning of human language — with its combination of meaningless phonemes into words, and of meaningful words into meaningful sentences (compositional semantics) — is unique in Nature. From a computational point of view, the most likely obstacle in the evolution of compositional semantics has been the necessity to perform operations on phonetic form and semantic structure in synchrony, perhaps requiring dedicated neural pathways. The most likely driving force for its evolution has been a selection pressure for an expressive, robust and learnable communication system under circumstances for learning and communicating with noise and time pressure.

5.4 Hierarchical Structure, Syntax and Recursion

Even simple utterances in natural languages go far beyond the rudimentary compositionality of the Campbell monkeys. First of all, they are not limited to combining two elements to create a third; the result of one combinatory operation in languages is usually again the input for the next combinatory operation. Thus, in a sentence like ‘happy people sing’, we first combine the meanings of happy and people, and then combine the resulting compound with sing. Human languages thus show hierarchical compositionality.

Moreover, words and phrases come in different categories. 'Happy' is an adjective that can modify the noun 'people'; combined they form a noun phrase that can be the argument of the verb 'sing'. Importantly, the syntactic categories of words and phrases, that determine what can be combined with what, are not always predictable from their semantics. Not only can we assess the grammaticality or ungrammaticality of nonsensical sentences (such as Chomky's famous pair 'colorless green ideas sleep furiously' vs. 'furiously sleep ideas green colorless', Chomsky, 1957), but syntactic constraints can also make sentences impossible that would semantically work perfectly well ('*the asleep child', '*John sang the Marseillaise his heart out', Culicover et al., 2004). Thus, natural languages employ a system of syntactic constraints that functions, at least in part, independently from semantics. The parts of a sentence over which syntactic constraints are defined are called phrases or constituents. Also for such syntactic phrases it is true that they do not always correspond one-to-one to semantic units; the hierarchical structure in the syntactic domain is called hierarchical phrase-structure.

Finally, in natural language sentences we can observe that a phrase of one particular syntactic category can be embedded in a phrase of the same syntactic category. Thus, a phrase like 'the man on the moon' is a noun phrase, but embedded in it we find another noun phrase: 'the moon'. A sentence like 'Luggage people leave behind is destroyed', contains 'people leave behind x' that linguists analyse as being of category sentence. This property of language is called recursion. In the debates on language and evolution a further distinction has played a key role: if the embedded phrase always ends up on the right or left edge of the larger phrase (as in the first example) this is called tail recursion; if phrases get nested in the middle of the larger phrase it is called center-embedding (as in the second example).

There is much disagreement in linguistics about the exact nature of hierarchical compositionality, phrase-structure, syntax and recursion, but there is no doubt that human languages show patterns that invite descriptions in these terms. In animal communication systems, in contrast, there is very little that comes close. In some song bird species, the song repertoires invite a description in terms of so-called finite-state machines (or hidden markov models): many songs here share a similar overall structure, but, for particular parts of a song, they differ in the number of repetitions of one or more elements or the choice for one variant or another (e.g., Okanoya, 2004). Although birdsong researchers describe this as 'song syntax', it's clear that it's very different from language: there is no semantics that the syntax can be independent of, and there is no real sense in which the system is recursive (let alone exhibiting center-embedding).

In humpback whale song researchers have also discovered relatively complex structure. Researchers describe the songs as being built up from themes, consisting of phrases, consisting of units, in turn built up from subunits. Hence, whale song might rightly be characterized as hierarchical (and a similar case can be made, though less pronounced, for many bird song species). However, there is no reason to assume a compositional semantics for these songs or a recursive structure. Also Suzuki et al.'s (2006) sophisticated analysis of humpback song does not establish

the need for a descriptions in terms of center-embedding, even if it does reinforce the conclusion that the songs are hierarchical.

Finally, a strong animal contender for the ability to process center-embedded, hierarchical structure is the bonobo Kanzi, who was exposed to human language from birth. Kanzi has been at the center of a long standing controversy about the language-abilities of apes. Unfortunately, the facts about what Kanzi could and could not do are hard to obtain. One side, represented by lead researcher Sue Savage-Rumbaugh, has tried to make the case for very advanced abililities (e.g., Savage-Rumbaugh & Lewin, 1994), but much of the presented evidence consists of video footage (which lacks crucial statistical information) or experimental data from designs that aren't up to today's standards in the behavioral sciences. The other side has often been dismissive without access to the relevant data and seems to have been driven in part by preconceptions about an innate language faculty (e.g., Pinker, 1994).

An interesting exception to this state of affairs is a recent paper by Rob Truswell (in prep), who reanalyzed a database composed by Savage-Rumbaugh and co-workers of spoken instructions to Kanzi and his responses. Truswell finds that Kanzi's performance is impressive in general: the ape seems very well capable of combining the meanings of several words. However, in most of the sentences used in the database, a correct interpretation of the instruction is not dependent on sensitivity to the hierarchical structure of the sentence. Truswell identifies a class of sentences where this sensitivity is crucial (sentences with NP-coordination) and finds that Kanzi's performance on those sentences is at chance level. These are sentences like 'Kanzi — put the coke and the milk in the fridge'. Assuming that Kanzi knows the meanings of all content words and knows that fridges can't be put into coke or milk (which is indeed an impressive achievement already), there are four possibilities for what goes into the fridge: nothing, coke, milk or both. Averaging over all 18 such cases in the database, Truswell finds that Kanzi is only correct 22% of the time.

A quite different approach to comparing syntactic abilities between humans and other species is pioneered by Fitch & Hauser (2004). They tested Tamarin monkeys on their ability to detect particular patterns in sequences of syllables. When one group of Tamarins had heard sounds conforming to the pattern ABAB or ABABAB, they reacted with surprise when confronted with the sound patterns AABB or AAABBB. However, another group of Tamarins first heard AABB/AAABBB and then failed to notice the change to the other patterns. Because the second pattern is typically used in mathematical work on center-embedded recursion, Fitch & Hauser interpreted these results as showing that the monkeys were unable to process such center-embedding. Subsequent work has shown that starlings (Gentner et al., 2006) and zebrafinches (van Heijningen et al., 2009) can, like humans, learn to distinguish between the two types of patterns. However, although the earlier papers generated much debate about whether or not animals can process recursive structures, van Heijningen et al. argue that the experimental set-up used in the experiments is problematic and that none of the results so far has really demonstrated the ability or inability to process center-embedding in any species. Rather, the results of van Hei-

jningen et al. show that each of the zebrafinches in their experiment exploit one of many possible non-recursive strategies to successfully distinguish grammatical from nongrammatical stimuli, and that the statistical analysis from earlier papers, where results from mutiple subjects were averaged, fail to correctly control for these alternative explanations. Hence, although this type of study might become important in the future to answer comparative questions about grammar, current results are inconclusive and more research is needed on this issue.

In conclusion, the way humans combine meaningful words to form complex sentences, guided by a system of semantic and syntactic categories and rules (collectively labeled grammar), is unique in nature. The computational complexity of this behavior, the absence of anything similar in animal communication, the failure of extremely intelligent apes to master it, and the fact that it makes language an extremely powerful system, together make a strong case that there is a true adaptation at play here.

Interestingly, for many of the claimed unique design features of language, the uniqueness seems to depend on the presence of grammar: words are symbols because of grammar, words might be learned efficiently because of grammar, talking about things remote in space and time (Hockett's displacement) is possible because of grammar, and human compositional semantics differs fundamentally from that of bees and Campbell monkeys because of grammar. Hierarchical phrase structure and the possibility of recursion and center-embedding, follow, it seems to me, from the way grammar allows us to combine words. The core component of grammar is hierarchical compositionality (other components are the syntactic constraints that are independent from semantics, but these are less crucial from communication and plausibly the result of preexisting idiosyncracies of the human brain); hence, hierarchical compositionality is at the top of the list of candidate features that make human language unique.

6 Towards an evolutionary scenario

6.1 Evolutionary scenarios: why and how?

So far, I have reviewed a number of traits of humans that seem directly involved in speech and language, and enquired to what extent they are shared with other animals. This exercise has led me to identify a number of candidate adaptations, some of which seem essential for a spoken, complex language to have emerged at all (vocal learning, vocal control, grammar), whereas others are more likely to be consequences of the new selection pressures that the use of a spoken language brought (optimized vocal tract shape, loss of air sacs). How do those fit into an evolutionary scenario that explains why humans and not other species have language?

By formulating a specific scenario I risk being accused of entering the realm of speculation, as so many theories on language evolution did before. However, as

long as we emphasize the hypothetical nature of any favoured scenario, I don't think much harm is done. Moreover, complete scenarios are in fact necessary if we want to investigate the relation between various proposed adaptations and evaluate the plausibility of each step in the context of the other steps. Knowing the place in a particular scenario further helps to focus our attention on the relevant evolutionary innovations, and evaluate their likelihood using modelling and the detailed analysis of data where available. Scenario building is thus actually necessary to move beyond speculation.

For evaluating the plausibility of various evolutionary scenarios that account for the comparative data discussed in this chapter, we can turn to various other fields, including comparative psychology for data on non-linguistic behavioral differences between humans and other primates and paleoanthropology for data on the evolutionary history of the human species. Additionally, evolutionary theory sets constraints on such scenarios, in particular by clarifying which components a scenario must involve. This is not the place to review the many findings from these various fields, but a few observations are useful to decide on what shape our scenario should have.

First, there is a whole suite of behavioral or cognitive abilities, other than language, that make humans stand out among animals, including advanced reasoning, consciousness, music, social cognition and theory of mind (knowing about the thoughts of others), the ability to imitate movements and sounds and our willingness to cooperate and share resources and knowledge. Together with uniquely human features of our life history (long helpless period in infancy, delayed sexual maturity, long post-reproductive life) and anatomy (reduced hair cover, sweat glands and upright posture), some researchers speak of a complete package of 'humanness' (e.g., Jones et al., 1992). It is clear that, a priori, an evolutionary scenario that assumes a common cause, or several common causes, for all of these different aspects of humanness is more plausible than an evolutionary scenario that assumes a distinct selection regime and evolutionary adaptation for each of them separably.

Second, from genetic and archeological data we know that the last common ancestor of chimpanzees and humans lived about 7 million years ago. A whole range of hominin species has been identified from fossil findings, ranging from the more ape-like Australopithecus afarensis closer to that common ancestor, to the much more recent Homo heidelbergensis occurring just before the appearance of anatomically modern humans about 200,000 years ago. One thing that is striking about those 7 million years is that most of that period involved only very slow changes. For instance, from about 2.6 million years ago hominins used simple stone tools, which remained virtually the same for a million years until hand axe technology first appeared in Homo ergaster. Then, in the last 100,000 years developments start to pick up speed. Art appears 80-100,000 years ago, modern humans spread around the globe (including the Americas about 12,000 years ago), agriculture is invented about 10,000 years ago, writing about 7,000 years ago and human history took off from there. A key factor in judging the plausibility of an evolutionary scenario of humanness, including language, is whether it can account for such a sudden speed-up in the evolutionary development.

Third, an evolutionary scenario describes a sequence of innovations, and evolutionary theory tells us to consider, at each step, whether the variation required for selection to operate would have been present and whether selection would have favored the proposed innovations among the many other possibilities. Focusing on the role of selection, there are two major obstacles in scenarios of language evolution. The first is that selection for linguistic traits is typically frequency dependent: the advantages of a trait usually depend on how many other people in a population already have it. For instance, knowing a particular word or grammatical construction is of little use if no-one else is able to understand it. As novel traits are initially always rare (because innovations in biological evolution are generated by rare mutations), this creates a kind of catch-22 situation: each innovation, even if it represents a true improvements when adopted, is initially selected against and therefore never becomes abundant enough to start conveying its advantage and thus be selected for. The second major obstacle can be called the 'problem of cooperation': linguistic innovations that improve the efficiency of information transfer are often not in the interest of the speaker, but only in that of the hearer. Hence, although not impossible, it is difficult to see why evolution would lead to speakers to adopting it.

Both obstacles thus have to do with the fact that language is a social phenomenon. Both can be overcome in various ways, for instance through the mechanism called 'kin selection': if an individual interacts preferentially with other individuals that are genetically closely related (e.g., one's brothers or sisters), natural selection can under some particular conditions favor the evolution of altruistic traits. For the various steps in any proposed scenario, we need to check, as well as we can, whether these conditions are met.

For the plausibility of any proposed scenarios, this means that those that involve very many genetically specified linguistic innovations under social selection pressures, are a priori less likely. This is the case, for instance, of the scenarios proposed by Jackendoff and Pinker: to overcome the discussed obstacles these scenarios need language-external circumstances for millions of years to be continuously unusually favourable. Moreover, during those same millions of years none of those human- and language specific tricks were selected for in other great apes. More probable scenarios, in contrast, involve a positive feedback mechanism: a mechanism where the emergence of a rudimentary form of language fundamentally changes the evolutionary dynamics and makes selection for further linguistic traits more probable. In such a scenario, favorable circumstances during a shorter stretch of human evolution could have provided the seed for a self-enforcing process leading to full-blown language.

To be sure, these arguments do not establish to correctness of falsehood of any scenario, but only establish that, before we have considered any data on the biology of language, scenarios are a priori more likely if they involve common causes, explain the speed-up and do not involve too many population-dependent genetic innovations.

6.2 A scenario of the evolution of the cultural phenomenon 'language'

Combining these desiderata with the comparative evidence from sections 4 and 5, I arrive at the following scenario — hypothetical, but more plausible in my assessment than the alternatives and worthy as a working hypothesis.

The scenario starts out with the traits of the last common ancestor (LCA) with chimpanzees. Given the comparative evidence, I assume that the LCA, like modern apes, had an ability to handle hierarchical, conceptual structure in reasoning about the physical world, in reasoning about the behaviour of conspecifics and other animals (prey, predators, competitors) and in making plans. I assume the LCA had, like modern apes, a relatively rich communication system, with tens of vocal and gestural signals, that involved some learning (especially on the receiving side) but no true vocal learning and no compositionality. I assume the LCA, like modern apes, lived in groups with a limited form of cooperativity and at least a minimal degree of social cognition. Finally, I assume it had a complex brain, with quite advanced cognitive abilities compared to other mammals, well adapted for survival in its contemporary environment but also with a 'hidden potential' to develop even more complex cognitive skills under the right circumstances.

The first step is a process of biological evolution after the split of the chimpanzee and human lineage, adapting the hominin species to function in larger social groups, probably as a result of moving from the forest to savannah environments (a change of niche also thought to be involved in the evolution of bipedalism, sweating and running skills). Selection pressures for surviving in a group and as a group (with the typical mix of selectional mechanisms studied in social evolution theory, including kin selection, altruistic punishment and knowledge-for-status; see, e.g., West et al., 2007) then led to increases in social intelligence, in cooperativity, in the willingness to share information and in the size of signal repertoires. The need for larger signal repertoires, in turn, led to the increased reliance on learned vocalizations (vocal learning, gestural imitation), with learned, conventional meanings and combinatorial phonology (reuse of articulatory programs, but no compositional semantics).

With the appearance of a learned signal system, the circumstances were ready for the second step: cultural evolution kicked in, and the signals adapted culturally to pre-existing biases of the hominin brain, ears, hands and mouth. Because of the cultural adaptation, the communication system could become more complex than it could have become otherwise. I take the highest achievements of any non-human primate today as an estimate of what these hominins could achieve: large repertoires of conventional, arbitrary signals (vocal and gestural) and a rudimentary form of compositional semantics.

Step three is that once this communication system, due to cultural evolution, had started to form such an important aspect of life, it also started to change the course of biological evolution. The rudimentary language served as a medium to transmit knowledge from generation to generation, for instance by learning about food sources or relatively rare but grave dangers. Hence, language made those individu-

als that mastered it well more knowledgeable than their less talkative competitors, and thus more likely to survive and more attractive to mate with. More and more complex language thus led to more complex cognition and to increased biological selection pressure for both general cognitive and specific linguistic abilities, including those subserving speech such as vocal learning, vocal control and acoustic range. Moreover, proficient language users were likely to seek each other's company, and thus profit from their advanced abilities even in populations where those were rare. This provides a positive feedback mechanism through which the presence of language makes overcoming the obstacles from social selection pressures more likely.

In that run-away evolutionary process, I assume a kind of arms race between language users emerged in which at some point, step four, the ability for hierarchical compositional structure emerged. It is difficult to say how much the primate brain had to change to allow for hierarchical compositionality without even the beginnings of an understanding of how it is implemented in modern human brains. However, its striking absence in animal communication and in the achievements of language-trained apes and birds, combined with the fact that many uniquely human linguistic traits seem linked to it, strongly suggest a biological basis. Also from a computational point of view, hierarchical compositionality is special as it requires dedicated computational mechanisms to perform operations in the signal domain and meaning domain in synchrony. I would speculate that a neural pathway for synchronizing preexisting combinatorial operations in the conceptual and motor domain (i.e., combinatorial phonology) was the crucial innovation.

Once hierarchical compositionality (HC) emerged, cultural evolution could take the languages spoken (or signed) to unprecedented levels of complexity in step five. Given the enormous diversity in languages spoken and signed today, and the fact that any human child can learn any of them, I suspect there is little biological specialization for language beyond HC. Symbolism, duality of patterning, phrase-structure, recursion are all potentially indirect consequences of HC, and I see the vast variety of intricate patterns in phonology, morphosyntax and pragmasemantics as likely to be the result of cultural evolution adapting to the pre-existing features of the human brain and body under communicative pressures.

The final and sixth component of this scenario concerns the impact that the discovery of complex language could have had on other aspects of cognition. It has often been proposed that language facilitates reasoning, planning, music, consciousness, social cognition and other cognitive domains, but equally often scientists have made proposals where the direction of influence is the other way around. Although with the current state of knowledge it would be unwise to claim much certainty on any position, I favour a language-first scenario. From an evolutionary point of view, the main argument in favour of language as the foundation for the rest is that it is the only of the uniquely human cognitive functions that plausibly plays a role in all the other functions and can plausibly have facilitated its own evolution through the positive feedback mechanism discussed above. Without positive feedback, it would remain a mystery why there are no non-human animal species that share at least some of those functions. E.g., scenarios in which general intelligence is the driving force need to postulate millions of years of selection for intelligence until the

various thresholds for language, music, consciousness etc. are met, while no other animal species was apparently under selection long enough for intelligence to reach even one of those thresholds.

7 Conclusions

A solid, scientific understanding of the evolutionary origins of language will remain elusive for some time. This means that the field of language evolution will continue to be an attractive domain for speculation and fantasizing. However, we need not (and, indeed, should not) accept this as a final verdict of the field. Solid comparative research and formal modelling, often inspired by more speculative theories, have led to many new findings on how aspects of natural language relate to animal abilities and under which circumstances biological and cultural evolution will favor particular changes in those abilities. Taken together, this evidence points to a central role for vocal and gestural imitation as the basis for cultural evolution, and to hierarchical compositionality, as the essential and uniquely human feature of language needed in definitions of symbolism, duality of patterning, phrase-structure and recursion.

The evidence also allows us to evaluate the relative plausibility of various scenarios. Although different researchers might reach different conclusions, this exercise leads me to conclude that a central focus for research in this field ought to be on those steps in the scenario for which there still is embarrassingly little empirical and modelling evidence: the neural basis of hierarchical compositionality, the feedback mechanism of language fostering its own evolution and the possible roles of language in not just influencing but facilitating consciousness, reasoning, planning and music.

References

Alemseged, Z., Spoor, F., Kimbel, W. H., Bobe, R., Geraads, D., Reed, D., & Wynn, J. G. (2006). A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature*, 443, 296–301.

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

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Bloom, P. (2004). Can a dog learn a word? *Science*, 304, 1605–1606.

Botha, R. (2008). Prehistoric shell beads as a window on language evolution. *Language and Communication*, 28, 197–212.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. New York, NY: Harper and Row.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

Crystal, D. (1997). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press, 2nd edition.

Culicover, P., Nowak, A., & Borkowski, W. (2004). Learning constructions and the theory of grammar. In *Proceedings of the 2004 Stanford Child Language Research Forum*.

Dave, A. & Margoliash, D. (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science*, 290, 812–816.

de Boer, B. (2009). Acoustic analysis of primate air sacs and their effect on vocalization. *The Journal of the Acoustical Society of America*, 126, 3329–3343.

de Boer, B. (2010). Modelling vocal anatomy's significant effect on speech. *Journal of Evolutionary Psychology*, 8, 351–366.

Deacon, T. W. (2000). Evolutionary perspectives on language and brain plasticity. *Journal of Communication Disorders*, 33, 273–290.

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.

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

Fitch, W. T. (2005). The evolution of language: A comparative review. *Biology and Philosophy*, 20, 193–230.

Fitch, W. T. & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303, 377–380.

Frishberg, N. (1975). Arbitrariness and iconicity: historical change in American Sign Language. *Language*, 51, 696–719.

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

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., & Scharff, C. (2004). FoxP2 expression in avian vocal learners and non-learners. *Journal of Neuroscience*, 24, 3164–3175.

Harnad, S. (2003). Symbol-grounding problem. In *Encylopedia of Cognitive Science*. Macmillan.

Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203, 88–96.

Janik, V. M. & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, 60, 1–11.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 1016, 749–777.

Jarvis, E. D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science*, 5, 5–14.

Jones, S., Martin, M., & Pilbeam, D. (Eds.) (1992). *The Cambridge Encyclopedia of Human Evolution*. Cambridge: Cambridge University Press.

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: Evidence for "fast mapping". *Science*, 304, 1682–1683.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., Hublin, J. J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A.,

& Pääbo, S. (2007). The derived FOXP2 variant of modern humans was shared with neandertals. *Current Biology*, 17, 1908–1912.

Lachlan, R. F. & Slater, P. J. (1999). The maintenance of vocal learning by gene–culture interaction: the cultural trap hypothesis. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266, 701–706.

Levelt, W. J. & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50, 239–269.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.

Lieberman, P. (1984). *The Biology and Evolution of Language*. Cambridge, MA: Harvard University Press.

Martinet, A. (1949). La double articulation linguistique. *Travaux du Cercle Linguistique de Copenhague*, 5, 30–37.

Martínez, I., Arsuaga, J.-L., Quam, R., Carretero, J.-M., Gracia, A., & Rodríguez, L. (2008). Human hyoid bones from the Middle Pleistocene site of the Sima de los Huesos (Sierra de Atapuerca, Spain). *Journal of Human Evolution*, 54, 118–124.

Martínez, I., Rosa, M., Arsuaga, J.-L., Jarabo, P., Quam, R., Lorenzo, C., Gracia, A., Carretero, J.-M., Bermúdez de Castro, J.-M., & Carbonell, E. (2004). Auditory capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain. *Proceedings of the National Academy of Sciences, USA*, 101, 9976–9981.

Nowak, M. A. & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences, USA*, 96, 8028–8033.

O'Grady, W. D. (2005). *How children learn language*. Cambridge: Cambridge University Press.

Ohms, V. R., Snelderwaard, P. C., ten Cate, C., & Beckers, G. J. L. (2010). Vocal tract articulation in zebra finches. *PLoS ONE*, 5, e11923.

Okanoya, K. (2004). Song syntax in bengalese finches: proximate and ultimate analyses. *Advances in the Study of Behavior*, 34, 297–346.

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior*, 7, 371–384.

Payne, R. S. & McVay, S. (1971). Songs of humpback whales. *Science*, 173, 585–597.

Pinker, S. (1994). *The language instinct: how the mind creates language*. New York, NY: Harper Perennial.

Pinker, S. & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, 95, 201–236.

Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, Phoca vitulina. *Canadian Journal of Zoology*, 63, 1050–1056.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society of London B*, 272, 941–947.

Riebel, K. & Slater, P. J. B. (2003). Temporal variation in male chaffinch song depends on the singer and the song type. *Behaviour*, 140, 269–288.

Savage-Rumbaugh, S. & Lewin, R. (1994). *Kanzi: the ape at the brink of the human mind*. New York, NY: Wiley.

Seyfarth, R. M. & Cheney, D. L. (1997). Some general features of vocal development in nonhuman primates. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development*, (pp. 249–273). Cambridge: Cambridge University Press.

Smith, E. C. & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, 439, 978–982.

Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. *Journal of the Acoustical Society of America*, 119, 1849–1866.

Tamariz, M. (2005). Configuring the phonological organization of the mental lexicon using syntactic and semantic information. In *Proceedings of the 27nd Annual Conference of the Cognitive Science Society*, (pp. 2145–2150).

Tomasello, M., Call, J., Warren, J., Frost, G., Carpenter, M., & Nagell, K. (1997). The ontogeny of chimpanzee gestural signals: A comparison across groups and generations. *Evolution of Communication*, 1, 223–259.

van Heijningen, C. A. A., de Visser, J., Zuidema, W., & ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences, USA*, 106, 20538–20543.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences, USA*, 92, 930–933.

von Frisch, K. (1974). Decoding the language of the bee. *Science*, 185, 663–668.

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, 20, 415–432.

Westermann, G. & Miranda, E. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language*, 89.

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63, 293–299.

Zuidema, W. (2003). Optimal communication in a noisy and heterogeneous environment. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in Artificial Life (Proceedings of the 7th European Conference on Artificial Life)*, (pp. 553–563). Berlin: Springer Verlag.

Zuidema, W. & de Boer, B. G. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, 37, 125–144.