

## Some modern considerations for thinking about language evolution: A discussion of *The Evolution of Language* by Tecumseh Fitch

Adam Kendon

*Institute for Research on Cognitive Science, University of Pennsylvania  
Philadelphia, PA*

### **A little history**

The question of how humans came to have language has been raised ever since humans first recognised themselves as languaging creatures, and the ability to language has always been seen to be the single most distinctive feature of humanness. No humans have ever been encountered who did not speak, but speaking has never been found in any other kind of creature. Nevertheless, it has always been recognised that other creatures had ways of communicating, especially by means of their voices, and many have sensed that, despite differences, human speaking was related to this. One ancient debate has been, then, whether or not we can accept continuities between animal and human expression, or whether there is an unbridgeable gulf.

In the modern era, from the middle of the eighteenth century until the end of the nineteenth century, inquiries into language origins were often undertaken. After this they ceased almost completely because, so it seemed, with the expansion of empirical knowledge, the gap between what we could reliably know that was relevant to the inquiry and what we would need to know to ground the inquiry on a solid base of observable facts had become glaringly apparent. Solutions to the problem of language origins, it was felt, could never be more than fairy tales, one person's views being as good as another's. Thus Dwight Whitney, writing in 1873, declared that "no theme in linguistic science is more often and voluminously treated...with less profitable result in proportion to the labour expended." He judged the "greater part of what is written upon this topic [of language origins]" to be "mere idle talk." (Whitney 1873-4, as quoted In Jespersen 1922: 412). As a result, the whole topic became disreputable. As it is common to note, in 1865 the Linguistic Society of Paris explicitly banned all submissions on this topic, and the London Philological Society followed suit a few years later.<sup>15</sup>

Notwithstanding this, from the late nineteenth century until the beginning of the period when the topic was to become fashionable again, there were scholars, linguists among them, who continued to contribute to the discussion. For example, important ideas were contributed by Charles Darwin and Otto Jespersen, as we shall see. However, these

---

<sup>15</sup> For a history of inquiry into language origins up to the end of the nineteenth century, see Stam (1976).

contributions remained isolated. They were neither a part of, nor did they initiate, any thread of discussion. The modern revival, which began in the United States in about 1960 and from which contemporary debate on the topic can be seen to have stemmed, was due to new initiatives by scholars who were taking into consideration information and ideas that had not been available before.

If one were to name a year to be the birth-year for the current language origins debate, one might well choose 1964. This was the year in which Charles Hockett, a well-established linguist reared in the then dominant structuralist tradition associated with Leonard Bloomfield, together with Richard Ascher, an anthropological palaeontologist, published an article in *Current Anthropology* entitled "The human revolution" (Hockett and Ascher 1964). In this article the authors made a bold attempt to bring together what was then known about hominid palaeontology, new understandings of the environmental changes in Africa, and speculations about the important consequences of bipedalism, to suggest what may have been involved in the evolutionary emergence of humans. However, almost for the first time within the framework of discussions of this sort, they presented the steps and stages involved in the evolution of language. As they pointed out, in previous discussions of human evolution from a palaeontological perspective, language was generally overlooked or dealt with purely in terms of evidence for the presence or absence of articulate speech. In this article, the authors make use of Hockett's notion that human language is a complex of "design features", some of which are found to be in common with other animal communication systems (see Hockett 1960). By setting out these features it was possible to specify more precisely what it was that had to evolve to bring into being a system of communication with all the features of human language. On this approach, language did not evolve as a single package, but in a more piecemeal fashion, each separable feature having its own evolutionary history. Hockett and Ascher supposed that an early step would have been for a primate call system - their proto-hominid model was based on what was then known of gibbon calls, which had been described with some thoroughness by C. R. Carpenter (1940) - to be transformed from a closed system to an open system through 'blending', a process by which calls of different meanings could be joined together to make calls with more complex meanings. It was then further transformed by the discovery of the possibility of what Hockett referred to as "duality of patterning", according to which meaningful units within the system are created through combinations and re-combinations of sound units that have no meaning in themselves but which function to keep meaningful units distinct from one another. Hockett and Ascher's article, in accordance with *Current Anthropology* practice, was accompanied by commentaries by a number of other scholars. In this way, this discussion of language origins was brought to the attention of a wider audience of scholars in disciplines such as palaeoanthropology, archaeology, anthropology, and even linguistics. The article had the effect, thus, of beginning the process of re-legitimizing the topic of language origins.

Shortly after this, and from another quarter, came the publication of Eric Lenneberg's *Biological Foundations of Language* (1967). In this book Lenneberg set out to show, in great detail, the biological features of humans that seemed to be special adaptations for speech and language, arguing that humans are biologically specialised as speaking creatures. This book contained an appendix by Noam Chomsky in which the idea that humans

are equipped with what Chomsky called a “Language Acquisition Device”. He suggested, that is, that humans have an innate device that shapes their linguistic productions, as they begin to become capable of them, according to pre-established grammatical patterns. He proposed the idea that humans have a “language organ” much as they also have a liver, a heart, and so forth.

Chomsky had, already, with his publications of 1957 (*Syntactic Structures*) and 1965 (*Aspects of the Theory of Syntax*), initiated what has often been regarded as a “revolution” in American linguistics, of which an important feature was the directing of linguists’ attention toward the *faculty* or *mental capacity* for language as being the true object of linguistic study. This idea, that language was to be understood as a cognitive operation, meant that linguistics was to be regarded as a branch of *mental* science. This firmly re-set the study of language within an organic, and therefore, biological framework. As a consequence, the issue of the biological origin of language was again relevant.

A next important development in the modern debate about language origins, was the announcement by Allen and Beatrice Gardner, in 1969, that they had successfully taught a chimpanzee to use, in an apparently symbolic manner, manual expressions borrowed from American Sign Language (Gardner and Gardner 1969). This created a considerable sensation. Chimpanzees had long been recognised as being very close, biologically to humans and it seemed that, with a little training, it ought to be possible to teach them to speak. Valiant efforts had been made by the Kelloggs’ (in the 1930s) and again by the Hayes’ (in the 1950s), who had tried the experiment of rearing a baby chimpanzee in their own home as if it was a member of their family (Kellogg and Kellogg 1933; Hayes 1951). Their efforts to teach the chimpanzee to speak in both cases ended in failure. Now, however, with the Gardners’ work, using the medium of manual action, a languaging ape seemed possible.

The achievement of the Gardners’ sparked much controversy, since this seemed to suggest that the great Rubicon separating man from beast that Max Müller had defined in 1873 and which he had declared no animal would ever cross, was now about to be crossed.<sup>16</sup> At the same time, it prompted a real effort to look into the true nature of sign language. Some years before, in 1960, William Stokoe had published what is generally taken to be the first attempt to analyse “a system of visual communication as used by the deaf” (as his publication was titled - see Stokoe 1960) using a framework derived from the so-called structuralist analysis of spoken languages, which showed that such a system in many ways had its own linguistic structure. Stokoe’s work aroused interest and gained acceptance rather slowly, however. It was not until the work of the Gardners gave urgency to the question of the linguistic status of sign language that work commanding wider notice was commenced. Thus the Gardners’ work prompted Ursula Bellugi to begin her project on sign language at the Salk Institute which led to the publication of *The Signs of Language* (1979). This played a very important role in demonstrating that these systems have all the characteristics needed to deserve the label (and therefore the prestige) of ‘language’ and in

---

<sup>16</sup> For an account of this see Radick (2008). For the controversy that the whole phenomenon of language-taught apes gave rise to, see Sebeok and Sebeok (1980) and Wallman (1992) among many other publications.

showing how interesting it would be to delve further into them.<sup>17</sup>

But the Gardners' achievement was also, of course, of great importance for the development of interest in the question of language origins. It was one of the main pieces of evidence that Gordon Hewes drew upon in his "Primate communication and the gestural origins of language" which he published in *Current Anthropology* in 1973. In this article, Hewes drew together a wide range of evidence, including findings from the then recent new work on the behaviour of chimpanzees in the wild, recent work on the neurological bases of primate call systems, as well as anthropological and historical findings, to argue that language first emerged in the gestural modality. He suggested that a form of gestural *protolanguage* was established, which then, subsequently, was transposed into a spoken form.

Hewes' article, followed as it was by extended comments by many scholars, had a greater impact than the earlier article of Hockett and Ascher. Perhaps this was because the Gardners' work had made the question of language origins more pressing. A major symposium on language origins held at the New York Academy of Sciences in September of 1975 then followed. This symposium, which resulted in a very large book (Harnard, Steklis and Lancaster 1975), involved the participation of many prominent scholars, including several linguists - Noam Chomsky among them. This further contributed to the serious attention that the topic of language origins was beginning to receive in academic circles. Thereafter, investigations extended in every direction: the neurology of speech, the nature of sign language, the study of human gesture, in palaeoanthropology, in linguistic theory, in the observational study of the behaviour of apes and monkeys in the wild.<sup>18</sup>

Further impetus was given to this discussion in 1990 by an article in *Behavioral and Brain Sciences* by Steven Pinker and Paul Bloom. This article accepted the theoretical framework advocated by Noam Chomsky. However, the authors argued that the human specialisation for learning a grammatical language could have evolved by way of a conventional neo-Darwinian process. The impetus that this discussion provided was part of what lay behind the subsequent organisation of a successful series of biannual international conferences on language origins, under the inspiration and guidance of Jim Hurford and Chris Knight, among others.<sup>19</sup> The series of volumes that

<sup>17</sup> Kendon (2002) summarises this history. See also Bronowski and Bellugi (1970) and Bellugi (1981).

<sup>18</sup> Although it has been claimed that this was the first symposium to be held on language origins sponsored by a major scientific academy since the essay competition promoted by the Berlin Academy of Sciences in 1769 (essays submitted in 1771, with Herder winning first prize), a symposium on this topic was held in 1972 in Toronto at the meetings of American Anthropological Association. This was published as Hewes, Stokoe and Wescott (1974) and it should not be overlooked. It was a result of this meeting that William Stokoe became interested in language origins (leading, ultimately, to his book of 2001, among other things). Furthermore, the publishing company Linstok Press, founded by William Stokoe and which was the first publisher of the pioneering journal *Sign Language Studies*, was apparently started as a way of publishing the 1972 symposium which, at the time, no other publisher was prepared to touch!

<sup>19</sup> The first of these conferences was held in Edinburgh in 1996. The most recent, known as *Evolang 8*, was held in 2010 in Utrecht. Volumes published from these conferences, or from related occasions, include Hurford, Studdert-Kennedy and Knight (1998), Knight, Studdert-Kennedy and Hurford (2000), Wray (2002), Botha and Knight (2009a, 2009b)

these conferences have produced, together with an ever increasing number of articles and books on the topic, testifies to a veritable explosion in interest.<sup>20</sup> It is as if, now, at last, there has developed a sense that an understanding that can lead to a solution is somehow within sight and, as a result, is worth struggling for.

However, anyone who gives even the briefest consideration to any of these more recent developments in the literature will discover that if anything intelligible is to be said on this topic it will be necessary to have an acquaintance with a very broad range of disciplines. Nowadays, whatever one's discipline, in beginning to explore this literature, one is likely to find oneself drowning in a vast sea of publications on topics that seem very far-flung from one another. One must read about bird song, fossils of the hyoid bone of early hominids, neuroimaging explorations of brain activity, sign languages, computer simulation studies of the development and change of artificial languages in artificial communities. It is rapidly becoming very difficult, if not impossible, for any one scholar to hold all of these diverse lines of investigation together and to arrive at an informed position.

Thus it is that the publication of *The Evolution of Language* by W. Tecumseh Fitch (Cambridge University Press, 2010) is to be greeted with both relief and enthusiasm. Although this is a long book (over 600 pages, altogether), it is written with such clarity and elegance that reading it is as pleasurable as it is informative. The book does us all a great service.<sup>21</sup> Fitch lays out almost all of the major components that any one would need who wishes to gain a grasp of how the problem of language evolution is currently being approached. As Fitch states (p. 3), his book is meant to provide “an overview of many different perspectives on language and the many types of data relevant to the debates, accepting each as a necessary component of some future synthesis.” In writing this book he says (p. 3) he has sought to fill the need for a “dispassionate survey of the available hypotheses and an even handed evaluation of their strengths and weaknesses in the light of currently available data.”

### **What Fitch's book contains**

This prospectus might lead one to expect a rather bland style, but just the opposite proves to be the case. Fitch is strongly committed to an approach that is fully informed by modern evolutionary biology. Especially important, for him, is the value of the comparative approach. As he explains, there are two aspects to this. On the one hand, by comparing species who are within the same descent grouping, or clade, as we do, for example, when we compare humans with chimpanzees or other great apes, and by identifying features that are homologous with one another in these various species, we can hope to reconstruct what traits characterised the common ancestor of the species within such a group. By means of this approach, Fitch attempts to reconstruct what features must have belonged to what he refers to as the Last Common Ancestor of humans and great apes, or LCA. Doing so allows us to be clear about what sorts of evolutionary transformations must have occurred to

<sup>20</sup> See, for example, Johansson (2005) who lists more than 2000 references, the majority of them dating from 2000 and after.

<sup>21</sup> Perhaps the book was finished in some haste. Both its Author Index and its Subject Index are woefully inadequate. This is unfortunate in a book which could be a useful reference work.

give rise to modern humans. On the other hand, it is also extremely useful to compare what appear to be parallel adaptations. For example, with regard to human speech, we find nothing which bears any resemblance to it within the group of species from which we reconstruct the LCA for apes and humans, although we do find certain cognitive capacities in the LCA that are also a part of human language. This means that speech is something that must have evolved after the divergence of the hominid line from the line that led to the great apes. On the other hand, in several different, unrelated vertebrate groups, including song birds, parrots, humming birds, bats, whales and dolphins, seals and elephants, we find capacities for more or less elaborate vocalisations, in some species used in a song-like manner, which are learned, rather than innate. Vocal learning and imitation are fundamental to human speech so, given these other vertebrate lines in which this is also found, it would be appropriate to study these to see what light this may throw on the kinds of evolutionary processes may be involved in the emergence of this characteristic. Indeed, the parallels between bird song and aspects of human speech have impressed investigators ever since Darwin drew attention to them in his *Descent of Man* (1871), and as a result there is a considerable body of literature investigating bird song, including its neurological basis (Catchpole and Slater 2008 is a recent survey).

Since this is a book about the evolution of language, Fitch includes in the first section of the book, two chapters which set out, first how language might be approached from a biological perspective, and then how language is to be defined. Here Fitch does not set up a categorical definition. Rather, he shows how we must think of language as a “suite of different but interrelated mechanisms” (p. 511). His approach is similar to that of Charles Hockett (Fitch acknowledges his debt to him) who, as we noted, proposed that language be regarded as an assemblage of “design features”. Fitch’s own list of features (given on p. 141) is rather different from those of Hockett, most notably, perhaps, because he includes pragmatics. Fitch notes the importance of using context to make inferences about meaning, the human capacity to take the other’s perspective and adjust his communicative acts accordingly, and what he calls *Mitteilungsbedürfnis*, a German word that refers to the urge people have to express their thoughts and share them with others. Fitch suggests that this is an important human drive, that must have played an important role in the development of language.

As Fitch argues, it seems likely that each of these design features or components of the suite that we refer to as ‘language’ has had a different evolutionary history. From this point of view, then, ‘language’ did not evolve as a single thing, but came about as a result of a particular combination of separately evolved traits or features. Thus we may expect rather different considerations are needed to throw light on the origin of speaking from those that might be needed to throw light upon how signals of any sort came to have semantic significance. The problem of the development of syntax may require different considerations again, while the subsequent development of languages as shared communication systems, and how this has shaped it and how it has shaped the niche or niches in which humans have come to live, and how this may have had consequences for our further evolution, these also are questions that demand somewhat separate treatments.<sup>22</sup>

<sup>22</sup> For an exposition niche construction theory in relation to language evolution see Odling-Smee and Laland (2009) and see also Bickerton (2009).

If we are to understand human language from the point of view of evolutionary biology, it is obviously important to compare human cognitive and communicative capacities with those of other species, not only looking at monkeys and apes, but much more widely. From the middle of the last century onwards there has been a great expansion of work on the cognitive and communicative capacities of other primates, of birds, as well as of certain other groups, such as wolves and dogs (Hauser 1996 is a recent survey). Appropriately, Fitch provides a long chapter which reviews much of this work. Here he shows how greatly our understanding of animal cognition and communication has changed. As a result of decades of careful observations of animals in the wild, often combined with some very sophisticated experimentation carried out in the contexts in which animals live, especially using the technique of 'playback' by which one can manipulate the calls and cries of conspecifics to see how they react to them, the conclusion, as Fitch summarises it, is that "animals possess a rich cognitive world, but are quite limited in their ability to communicate their thoughts to others." (p. 201).

Accordingly, it is clear that the common ancestor of humans and apes must have had a rich cognitive life and must have had a sophisticated capacity to interpret the vocal signals of its conspecifics and probably those of other species as well. However, it must have been much more limited in its capacity to engage in communicative acts that could serve in wilfully informative ways. According to current evolutionary theory regarding animal communication, however, there are many difficulties in accounting for how a system of cheap to produce, trustworthily informative communication signals that is characteristic of human language could have evolved. Very particular circumstances that created the development of co-ordinated attention and honest communication must have come to prevail in human evolutionary history to make possible the emergence of the sort of system language is. As Fitch says (p. 202), how this came about is a "central question for theories of language evolution." In a later chapter, Fitch offers his own solution to this problem in terms of a theory in which communication systems can be shaped by kin selection. We will return to this below.

The two chapters on language and the chapter on animal cognition and communication are included in Part I of the book which is titled "The lay of the land: an overview of disciplines and data relevant to language evolution." In addition to the three chapters discussed above, this part also includes a chapter on contemporary evolutionary theory. Here basic notions such as natural selection, sexual selection, kin selection and inclusive fitness are explained, fundamental concepts in current molecular genetics are expounded, and recent developments presented, such as how developmental phenomena are to be understood in evolutionary terms (so-called "evo-devo"). This chapter will prove useful to anyone needing a quick refresher on these issues, and it is useful to refer back to it when reading later sections of the book, which often make use of these ideas.

Part II of the book continues with the presentation of data and theories that are needed for developing an evolutionary approach to language, but now concentrating more specifically on the group of organisms that include humans. After a whirlwind tour of biological evolution from the beginnings of life to the emergence of the common ancestor of apes and humans in the Oligocene (from about 20 million years ago onwards), there is a chapter that

attempts to describe the features of the Last Common Ancestor of apes and humans (referred to as the LCA). This is followed by a chapter which surveys the current state of hominid palaeontology and archaeology.

With Parts I and II of the book, thus, Fitch has sought to equip the reader with much of the background that is needed to tackle some of the specific issues in the study of language evolution. Now, in Parts III and IV, Fitch turns to a detailed discussion of issues that are specific to the evolution of language. Part III contains three chapters devoted to the evolution of speech. In Part IV we find a detailed discussion of each of three different 'phylogenetic models' of language evolution: models which propose a lexical protolanguage, such as those of Derek Bickerton or Ray Jackendoff (Chapter 12), gesture protolanguage models, such as those proposed by Gordon Hewes and Michael Arbib (Chapter 13), and "musical protolanguage" (Chapter 14), according to which speech first emerged as a kind of song-like vocal communication system, acquiring the segmental character and semantic functions of modern speech only latterly. This theory, originally elaborated by Charles Darwin, subsequently developed by Otto Jespersen, and argued for more recently by Brown, Mithen and Falk, among current writers, Fitch sees as being in some ways the most promising of all.

The final chapter of the book is a summary and an evaluation of future prospects. Fitch concludes that all of the solutions that have been proposed to the various specific problems of language evolution have something to offer, but that no single author has been able to offer, up to now, a unified view that is consistent with all available data. Nevertheless, he concludes, "an empirical, hypothesis testing approach, embracing a comparative multi-component view, offers realistic hopes for real scientific progress in the next twenty years or so." (p. 512).

In what follows, I will treat in a little more detail what emerges from Fitch's discussion of the evolution of speech, and the three 'phylogenetic models' of language origins.

### **The evolution of speech**

Regarding the evolution of speech, Fitch concludes, first of all, that little can be derived about the origin of the speech apparatus from the comparative study of fossils. Fitch shows that even if it were possible to reconstruct the vocal tract of Neanderthals or even *Homo erectus*, as some have tried to do, this would tell us little about the speech capacities of these creatures. Philip Lieberman had attempted to reconstruct the vocal tract of Neanderthals and had argued that because the larynx in this kind of human appeared to be set very high, the Neanderthal would have lacked the pharyngeal space that is characteristic of the modern adult human, when, soon after birth, the larynx descends to a lowered position. Fitch argues, on the basis of analyses of the acoustic capacities of the vocal tracts of a number of different animals, including the tiger, the seal and the red deer, as well as the chimpanzee, that, contrary to what has been believed before, these animals would be capable of producing a human range of frequencies. Furthermore, he cites recent work in which the actions of the vocal apparatus are observed during vocalisation by using x-ray cine-photographic techniques which show that the position of the larynx at rest in animals such as the red deer is quite different from its position during vocalisation, when it can be



lowered considerably, thereby greatly altering the shape of the vocal tract. It thus remains quite possible that this kind of flexibility in the position of the larynx also occurred in the Neanderthal, something which cannot be gathered from the study of fossils. The only piece of fossil evidence that, according to Fitch, might bear on whether speech might have been present or not has to do with the size of the spinal canal in the thoracic section of vertebral column. In modern humans and Neanderthals this is larger than it is in other primates, and this could have to do with the more extensive innervation necessary for the muscles involved in breathing, the control of which is more elaborate because of the fine vocal control characteristic of speech. Evidently, expansion of the thoracic spinal canal began sometime in the million year period before the emergence of *Homo ergaster*, and before Neanderthals. This could suggest that in this period there was a development of increased vocal control. This could be compatible with the emergence of speech or with the emergence of some form of elaboration of vocalisation that could be a precursor to speech. Nevertheless, Fitch's conclusion from this discussion of the comparative fossil evidence is that little can be said about the timing of the emergence of human speech.

From his discussion of comparative studies of the vocal tracts of non-human primates and other mammals Fitch concludes that it does not seem to be the anatomical form, or indeed the articulatory apparatus of tongue, lips and larynx that marks out humans as specially capable of speech production. Rather, what is special lies in how vocalisation is neurally controlled. Chapter 9 provides a detailed review of recent work on this matter. Drawing upon work by Jurgens and his colleagues, especially,<sup>23</sup> It is concluded that there are direct cortical connections between the motor cortex and the motor neurons that control the muscles of the tongue and larynx. In most mammals, it seems, including non-human primates, the connections between motor cortex and the muscles of tongue and larynx are mediated by interneurons in the brainstem. In humans, the connections are direct. This may underlie the ability that humans show, that other primates do not, of being able to coordinate in a voluntary manner, actions of the vocal folds in the larynx with actions of the tongue and lips. It is perhaps this that makes possible the voluntary use of the voice in vocal learning and imitation, so fundamental for the capacity for speech.

Song birds, who are capable of vocal learning and imitation and of course produce very complex vocalizations, also have direct connexions between the frontal motor cortex and the syrinx (the apparatus that produces sound in birds). This is certainly a most interesting parallel. However, birds are quite different from mammals and parallels with them can only be indirectly illuminating for understanding the development of human vocal capacities. There is, however, a mammal that is capable of complex vocal imitation, and that is the seal. Fitch suggests it would be very worthwhile to investigate the neural connections of the vocal apparatus in this animal for this could shed a useful light on the evolution of this aspect of human vocal capacity which is crucial for speech.

Fitch concludes the chapter on the neural basis of speech with a discussion of the molecular genetic basis of motor control of vocalisation. Here we find a discussion of the FOXP2 gene that has attracted so much attention in recent years. As has been widely reported, various members of a family living in London were found to suffer from certain defects in speech production and it was found that these members had a deleterious

<sup>23</sup> See, for example, Jurgens (1995, 2000, 2002)

mutation in a regulatory gene known as FOXP2. This suggested that this gene plays a role in the development of fine motor control of the musculature involved in vocalisation. Comparative studies have shown that the FOXP2 gene is highly conserved and found in all vertebrates, but its human form differs in a quite specific ways from the form of this gene found in chimpanzees. Recently it has been shown that the human form of FOXP2 is also found in Neanderthals, so if the specific form that it has is indeed connected to a capacity for speech, this was present at least 400,000 years ago. What is intriguing is that it has been shown that the FOXP2 in song birds is also involved in the development of fine vocal control. As Fitch stresses, the FOXP2 gene is not a “language gene.” Its discovery, however, is the first time a genetic regulator has been found that affects the fine motor control of the vocal apparatus, which is crucial for speech. That it seems to play a similar role in song birds only strengthens the interest of the apparent parallels between song in song birds and speech in humans.

Whatever changes may have come about in neural organization that made speech possible in humans, these need not have been all that radical. In Chapter 10 Fitch summarises and synthesises the work of several investigators (including Jurgens, Philip Lieberman, MacNeilage and Deacon) in which separate attempts have been made to present syntheses of what is involved in speech production and its possible evolution. From this it emerges that whereas the lip, jaw and tongue movements crucial to the production of the acoustic variations characteristic of speech are all present and under voluntary control in chimpanzees, where they are employed in the processing of food and in certain other manipulatory activities, as well as certain non-vocalised, intimate kinds of communicative activity, as in lip smacking and grooming, what is novel in humans is the ability to coordinate these voluntary movements with voluntary control of the vocal folds in the larynx, controlling phonation and pitch.

This means that the behavioural system that enables humans to produce the complex of actions we term *speech* is a newly developed coordinative use of an already existing, evolutionarily more ancient system. As Fitch puts it (p.371), “speech has been ‘tinkered’ together from old parts”. This is, of course, utterly characteristic of biological evolution. Changes occur in the re-assembling and re-coordination of already existing systems and rarely involve the emergence of something completely novel.

As Fitch makes clear, speech, although fundamental for language, is not to be identified with it. An account of the evolution of speech is not the same as an account of language evolution. So it is that in Part IV phylogenetic models of language, not just speech, are discussed. After a brief historical chapter in which Fitch describes some of the earlier contributions, such as those of Condillac, Monboddo, Rousseau and Herder from the eighteenth century, and Darwin’s theory of language origins and Max Müller’s attack on this in the nineteenth century, Fitch presents three chapters in which, respectively, he considers theories about lexical protolanguages, the idea that protolanguage first emerged as a form of gestural expression, and finally the idea that language emerged as a differentiation from a system of vocal expression which had many of the characteristics of song.

### Lexical protolanguages

Chapter 12 begins with a review of the idea of ‘protolanguage’, as this has been developed by Derek Bickerton (e.g. Bickerton 1990, 1998) and subsequently, in a slightly different way, by Ray Jackendoff (e.g. Jackendoff 1999). According to Fitch, proponents of protolanguage all agree that pre-linguistic hominids and other primates, indeed, mammals generally, all were capable of entertaining conceptual models of their worlds. Protolanguage emerged first as a capacity to match signals *referentially* to these concepts. It is important to note that this is not exemplified, as some have supposed, by the so-called “functional referential” signals of primate alarm calls, such as those made famous by Cheney and Seyfarth in their study of vervet monkeys (Cheney and Seyfarth 1990), for in such cases these alarm calls, though differentiated according to type of predator, are not used in different contexts to *refer* to a given predator. Rather, they serve as a way of alerting fellow monkeys to different kinds of danger, when that danger is present. Exactly how functionally referential signals came to be *conceptually* referential, and so become units in a protolanguage is not anywhere clearly explained, however, and this does remain one of the unsolved mysteries of language origins.

Bickerton’s original concept of protolanguage was, essentially, that the first hominids to begin to acquire a protolanguage would have developed a lexicon, but there would be no syntactic organisation to their utterances. Bickerton proposed as models for protolanguage the asyntactic utterances of language trained apes and the earliest stages of language acquisition by children, where single words are used but without being a part of any grammatical sentence. He argues that the transition to syntax was sudden, and came about as a consequence of a mutation induced re-organisation of neural structures in the brain. As Fitch points out, apart from the rather uncertain value of the data from language trained apes, the actual way in which children appear to acquire syntax seems gradual and does not seem to support Bickerton’s idea. Better support for it comes from studies of transitions from pidgins to creoles. Bickerton himself first came to prominence in 1981 with the publication of his book *The Roots of Language*, which argued that the seemingly abrupt way in which children, in refashioning a language as their own, creating a creole out of the language materials of the pidgin which their adults spoke, suggested an innate component to the structuring of language which Bickerton called a *bioprogram* (Bickerton 1981). This idea has strongly informed the interpretation of the findings of Kegl, Senghas and colleagues in their study of the emergence of Nicaraguan sign language. This developed as an apparently new language when very young deaf children appropriated elements of what was a mixture of home sign systems, which they encountered when they entered a school for the deaf that had not itself been in existence long enough to establish a stable shared sign language of its own (Kegl, Senghas and Coppola 1999).

Jackendoff (1999) presents a modified and extended version of Bickerton’s idea of a protolanguage. Unlike Bickerton, Jackendoff proposes that it is possible to identify intermediate stages between the grammarless lexicon of a protolanguage and the emergence of fully grammatical speech. He also recognizes the importance of the role of the combinatorial nature of phonology in the generation of syllables which, for him, being combineable in multiple ways, provides the basis for the development of an unlimited pool of words.

Taking these two models of protolanguage together, Fitch then goes on to show that current evolutionary theory poses serious problems for the idea of a protolanguage, at least as it has been formulated hitherto. The basic problem is this. The assumption that all the proponents of protolanguage rely upon, indeed the assumption relied upon by so many who discuss language evolution, is that it would be obviously advantageous for hominids to communicate with one another, sharing information. However, as has become clear, an evolutionary explanation for the emergence of cooperative information sharing is very hard to formulate. This was a problem that troubled Darwin himself, and it is only in fairly recent years that a plausible solution has been arrived at. It is a problem because although a given *group* might benefit in comparison to another if more and better information was shared among its members, selection does not work at the group level. It works at the level of individuals, and why does it benefit an individual to share valuable information with another? It would usually be better for an individual to give misleading information to others, for then it would be able, for example, to take advantage of food sources without letting others know about them. As Fitch says, the free sharing of propositional information that language makes possible is one of the “central oddities of our species from an evolutionary viewpoint; one that cries out for selective explanation.” (p. 425).

The solution that Fitch offers, which has become widely accepted elsewhere in biology but has not, according to him, received much attention in discussions of language evolution, is to apply the theory of *kin selection*. From the viewpoint of current evolutionary theory, to share information among close relatives, especially dependent offspring, gives selective advantage to all those who share the same genes. So that, for selfish genes, for a mother to help her children is advantageous because this contributes to the survival of the shared genetic makeup. Once this point is accepted, the evolutionary mystery of information sharing among honeybees, for example, is cleared away - for all worker members of a beehive are sisters of one another. There is now a considerable accumulation of observations that suggest that animals related to one another do share information cooperatively if they are related, but do not if they are not. To take just one example, among ground squirrels, when outside their burrows a ground squirrel will emit alarm calls when it sees a predator, but only if it is in the company of kin-related individuals. A male ground squirrel who has recently joined a group as a mate, will not emit alarm calls at all, unless his own offspring are present. Similarly, vervet monkeys, famous for their differentiated alarm calls, do not emit alarm calls when they are among unrelated individuals. There is good reason to suppose, thus, that the honest sharing of information can develop among kin groups, when it does not among unrelated individuals. Emitting an alarm call, while useful to others, makes the individual who emits the call more vulnerable to predation, but this will not disadvantage its genes if, in emitting the alarm, it thereby reduces the vulnerability of its relatives and thus improves the inclusive fitness of those who share these genes. As Richard Dawkins made clear in his *Selfish Gene* (Dawkins 1976), it is the differential survival of genes, not of specific individuals, that counts in evolution.

Fitch argues, thus, that *kin selection* is the process that most likely made it possible for reliable information sharing to develop among primates and he emphasises the importance of this for understanding how a communication system like language could have come

about. He suggests that it was the advantage gained, particularly by the highly dependent young, of reliable information about the environment, how to obtain food, the kinds of dangers, what to avoid, and so forth, that would have been selected for, for in this way whole groups of related offspring could have been advantaged. He suggests that the speed and ease with which children learn language, and the way in which language exchanges develop between parents and offspring support this notion, a feature of human communication which, he suggests, is hard to explain without the idea of kin selection.

Fitch points out that the selective advantage of sharing information with related others can explain, or at least provide the circumstances in terms of which we can explain how the sharing of propositional information can have a selective advantage. It could thus provide an account of the circumstances in which the conveying of propositional information by communicative signals could emerge and thus lay the foundations for a protolanguage. To understand how, among humans, the use of this mode communication came to be extended to include *unrelated* individuals, on the other hand (and language is obviously used in this manner), other processes need to be invoked.

Accordingly, Fitch proposes a two stage model for the emergence of protolanguage and how it came to be a system that can be shared reliably beyond close kin. He suggests that the extension of the use of language beyond kin groups would not need any further biological changes. Once the propositional informational value of communicative acts came to be established, its further extension, “the implementation of regulated information exchange between adults” (as Fitch puts it) could “evolve culturally, as a set of social norms, without any further biological specialisations beyond those already present ‘for free’...” (p. 428).

This “second stage” of Fitch’s model deserves a good deal more elaboration. This is not provided. However, his discussion suggests that it would be interesting to examine further who, in fact, does share information with whom. If one reflects on this one can see that the development of the sharing of propositional information beyond circles of kin probably happened quite slowly and would have been clearly related to the extent to which it became possible for humans to extend kin categories to non-kin persons. The classificatory kin systems still in use today among Australian Aborigines could provide an illustration of how this kind of extension could come about. Today, in “advanced” societies, information sharing can still be highly differentiated. People are quite selective in who they talk to and what they tell others and biases related to kinship can often be strong. A person who sees another as similar in background and culture is more likely to consider them trustworthy. If the rules of information sharing in groups were to be examined it would probably emerge that these would reflect patterns of kin relations or patterns of relationships which could have a kinship interpretation.

Now although it seems that Fitch’s emphasis on kin selection as having a crucial role in creating the circumstances in which a system of trustable low-cost signals conveying propositional information could have evolved and goes further than any other competing model in making this evolutionarily plausible, as Fitch himself recognises, at least two very important issues are left unresolved. The first is the origin of syntax. The second and, in some ways more fundamental issue, concerns the origin of the actual signals employed in the creation of utterances. What behavioural material, we might ask, came to be

fashioned for the units of the protolanguage and how did these units come to have the semantic significance that enabled them to function in a protolanguage?

Fitch's treatment of the problem of syntax is fairly brief. We have already mentioned Bickerton's idea that grammar somehow emerged in one fell swoop, as a result of a 're-wiring' of the brain brought about by some mutation. As we have mentioned, Fitch rejects this suggestion, and it does indeed seem very unlikely.<sup>24</sup> Fitch seems to prefer Jackendoff's stage-by-stage approach, but few details are given regarding how these stages came about. Fitch's discussion of the problem, however, does point the reader in the direction of two possible avenues that would warrant further exploration. Both of these suggest ways in which the emergence of syntax can be understood as developments from already existing biological systems. That is, Fitch does not see 'syntax' as being a somehow separate phenomenon, that emerges autonomously, but sees it as an elaboration and specialisation of already existing dispositions. These are, on the one hand, conceptual constructs - that is, constructs that arise from the basic way in which all vertebrates, not just humans, understand the world of things and events and how events may relate to one another. On the other hand, the complex hierarchal structures of extended sentences could be understood as deriving from the way in which extended sequences of motor control are organised. In skilled and goal directed action, as in tool making, feeding routines, hunting routines, and the like, a kind of syntactic organisation can be detected. Fitch agrees that important aspects of syntactic organization could be derived from this.

The second, and more fundamental problem is the origin of the units of action that serve as referential signals, and how they came to do so. As Fitch points out, all of the proposed protolanguage models, including the 'gossip' model of Robin Dunbar (e.g. Dunbar 1996), also discussed by Fitch, as well as his own idea of it, take for granted that 'meaningful' signals are already available. This leads to the two final chapters of the book: Chapter 13 which discusses the idea of a gestural protolanguage, and Chapter 14 which discusses what Fitch calls "musical protolanguage" - which is the idea that speech emerged from a form of vocalisation that was similar to a kind of singing.

### **Protolanguage as gesture**

In discussing the idea that protolanguage first emerged as a form of gesture, Fitch first reviews in some detail the proposal to this effect put forward by Gordon Hewes in 1973. We already mentioned this publication in the introductory part of this essay. Hewes' article of 1973 is a landmark publication and, in many ways, must be seen as initiating the modern conversation on language origins. Hewes, first of all, was impressed with the achievement of the Gardners in teaching a form of manual gesture to a chimpanzee. This was in such contrast to the great difficulties that had earlier been encountered in trying to teach chimpanzees to speak. This, together with the flexible way in which chimpanzees and other apes were already known to make use of gesture in communication, suggested to Hewes that the adoption of gesture as a medium for language would be the path of least resistance, biologically. Ape and monkey vocalisations were understood to be limited and stereotyped in their repertoires and these did not seem to offer themselves as a model from

<sup>24</sup> Bickerton (2009) no longer supports this idea.

which anything like spoken language could be derived. There was also neurological evidence that suggested that non-human primate vocalisations were not under voluntary control. In addition, Hewes pointed out that gesture could readily still be used by humans for communication (he cited reports of explorers encountering strange peoples in distant lands being able to engage in complex communication via the use of gesture) and he also referred to the existence of sign languages. Hewes recognised, however, that to propose a gestural origin for language immediately ran into the fact that modern languages are *spoken*, so he was faced with the challenge of proposing how and why there should have been a changeover to using speech for language.

He suggested a number of reasons that might account for this. These included the idea that speaking would be more convenient as a form of linguistic communication, such as that it could work in the dark or that it made it possible to talk and do things with the hands and the same time. He also suggested that a phonetic form of language would make large vocabularies possible and would provide a better basis for lexical items to be stored and retrieved from memory. He also was favourable toward the “tongue gesture” hypothesis put forward by Sir Richard Paget (Paget 1930), according to which the mouth, in making movements parallel to gestural movements by the hands, would, if these were accompanied by vocalisations, produce complex articulated sounds which could serve as acoustic indices of these movements. Paget’s observations, which led him to suggest mouth-hand synergies that could lead to a sort of mouth gesturing, does receive some modern support in recent work by Gentilucci and his colleagues (see Gentilucci and Dalla Volta 2007), and Rizzolatti and Sinigaglia (2006) have recently sought to revive this idea. Recent interest in so-called ‘echo-phonology’ in signers, in which the mouth sometimes moves in sympathy with hand movements might also be mentioned here (see Woll 2009). However, as Fitch concludes, none of these proposals by Hewes ultimately provide a plausible model for the selective pressures that might have led to a switch from the manual to the spoken modality for language or for how the transfer might have occurred.

Subsequent advocates of the ‘gesture first’ scenario for language origins include William Stokoe, who pioneered the modern linguistic study of sign languages, and his colleagues David Armstrong and Sherman Wilcox (see Armstrong Stokoe and Wilcox 1995), and Michael Corballis (2002). Fitch does not consider their contributions in any detail (although he does mention them), perhaps because (or so he seems to imply) to him their suggestions are in many ways similar to those originally made by Hewes. Although this is arguable, it should be pointed out that Stokoe and his colleagues, especially, were able to bring to bear on the argument a much more sophisticated view of sign language than was available to Hewes and their contribution, especially as expressed in the book they jointly published in 1995 and also the contribution of Stokoe’s last book (Stokoe 2001) deserves a much more ample treatment than Fitch gives it here.

Instead, Fitch discusses the more recent ideas of Michael Arbib, who has taken into consideration the findings on so-called *mirror neurons* (see Arbib 2005). Mirror neurons are neurons, first discovered in the motor cortex of the rhesus monkey, which are found to be active both when the monkey observes a grasping action being made by another monkey or by a human being and when the monkey itself performs a similar grasping action. It suggests that the observation of actions by others can produce patterns of neural activity that

are similar to those that occur when the action itself is performed, and this suggests a neural basis for the understanding of others. It can provide a basis for *parity* of communicative actions (in the theory of Rizzolatti and Arbib 1998), that is, for an action to mean the same thing for both actor and recipient, a necessary condition for communication. Arbib attaches much importance to the fact that these mirror neurons were found in a region of the macaque monkey's brain that is considered to be homologous to Broca's area in the human brain, an area known to be much involved in human speech. As first understood, it was proposed that the neural circuitry of the mirror system provided the foundation for the audio-motor mirror capacities which, presumably, are at the foundation of the human capacity to speak.

In its original formulation, the mirror system hypothesis was criticized on a number of grounds. In particular it was pointed out that rhesus macaques, and indeed other monkeys, do not imitate, and the function of mirror neurons in imitation, a process fundamental to the development of any sort of shared communication system such as language, is thus not clear. It has also been pointed out that although mirroring might allow the recognition of the grasping actions of others and, with some additions, allow for the other to imitate such actions, this does not provide the basis for how an action might acquire referential meaning. Furthermore, it is hard to see how the mirror system, as the basis for action recognition, can also be extended to the recognition of features of the environment, such as fruit and trees and predators, and how actions referring to these things could be derived from mirror recognition of actions.

Arbib has responded to some of these criticisms by elaborating a so-called "extended mirror system hypothesis". This is a complex hypothesis, which Fitch makes no attempt to summarise. However, there are two important points in Arbib's later work which deserve emphasis. First of all, one of his arguments for the idea of a gestural protolanguage is that by way of mimetic gestures it would be easier for the pre-linguistic hominid to "break through" the restricted repertoire of meanings possible in the vocal system than it would were the capacity for mimicry to extend first to vocalisations. This is because, according to Arbib, following principles of visual imitation, gesturing would offer so much more scope for representing objects and actions and their relationships than would seem possible if one relied only on an auditory medium. Second, unlike Hewes or Corballis, Arbib does not suppose that first there was a gesture protolanguage that later somehow switched over to a spoken language. Rather, he suggests that the transition from a gestural system augmented by vocalisations to a mainly vocal system was a gradual one. This he calls an "expanding spiral" and suggests, thus, a long period during which both systems were subject to selection, each "scaffolding" the other.

Fitch himself seems to be sympathetic to the idea that a gestural protolanguage could have provided a way in which a system of signals could develop an expanded semantics. As he puts it (p. 457), "gestures and mirror neurons appear to provide a good way *into* a meaningful protolanguage". However, even if one might agree to this (and neither Arbib nor Fitch provide any hint as to how an action of imitation ever comes to be *recognized* as such, and thus they offer no idea as to how it ever might acquire status as a *referentially semantic* action), one is still left without a convincing account of why



vocalisation took over as the principal medium of language.<sup>25</sup>

### Protolanguage as musical

So how *is* speech to be accounted for and why is it the dominant vehicle for language? That is, how did the elaborate control of the vocal apparatus that is characteristic of humans emerge and how did it acquire the capacity to serve as a *semantic* signalling system? In Chapter 14, the last main chapter of the book, Fitch elaborates on the idea that speech is a development of what he calls a “musical protolanguage” or, as he also calls it, a “prosodic protolanguage”. The relationship between speech and song has long been noted. It was discussed by Rousseau in the eighteenth century, and Charles Darwin, writing in 1871, supposed that there was an evolutionary relationship between song and speech. Later writers who have also proposed the idea include Otto Jespersen and, much more recently, Brown and Mithen.<sup>26</sup> Fitch’s presentation and discussion of the idea draws mainly on the work of Darwin and Jespersen.

As Fitch had made clear earlier in the book, and as he reiterates here, it is necessary to keep *speech* separate from *language*. Looked at in this way, one can view “phonology” (as he calls it here) as an autonomous generative system, according to which more or less unlimited sound sequences can be created through re-combinations of repertoires of “phonic elements”. These are not to be confused with phonemes, since, as Fitch points out, this is a concept employed in relation to a system that conveys linguistic meaning, and in prosodic or musical protolanguage we are not yet dealing with such meaning.

The justification for supposing such a system of “bare phonology” is based on a number of considerations.<sup>27</sup> First, when children are acquiring language, they go through a sort of practice phase of “bare phonology” when they babble. In this period the child seems to be trying out his capacities for articulation and exploring the range of sounds that it is possible to make, producing complex sequences of articulated vocalisations coordinated with intonation patterns - certainly a form of “pre-speech”, but with no linguistic meaning. This in itself suggests that “phonology” is a separate system.

In addition, adult speakers may often make use of meaningless syllabic sequences, simply for sonic and rhythmic effect, as in various kinds of singing (‘scat’ singing is one good example) and ritual performances. And even when we are dealing with linguistic speech, in a variety of contexts, much use may be made of nonsense words, repetitions, and the like, as in many varieties of song and poetry. Such phenomena have led a number of writers, as already mentioned, to propose that there are many parallels between speech and song, and indeed a survey of the ethnographic literature suggests that often the boundaries between what can be considered speech and what can be considered song are by no means

<sup>25</sup> See Kendon (2008, 2009, 2011) for further critical discussion of the ‘gesture protolanguage’ idea.

<sup>26</sup> See Brown (2000) and Mithen (2005) for details.

<sup>27</sup> Instead of “bare phonology” I suggest the term “lallation” adapting “-lalia”, a terminal element from a Greek word which means “talk” or “chatter”, used in English to refer to various kinds of speech disorder or strange practice (as in “echolalia”). Here, however, “lallation” is to mean the capacity to produce articulate sounds that are speech-like and can be used in talk, as in a baby’s babble or in nonsense talk or scat signing. I prefer it to Fitch’s “bare phonology” for then one does not have to apologise, as Fitch does, for using the word “phonology” in a sense different from its usual meaning in linguistics.

always clear cut.

One is led to the view, thus, that the capacity for producing, using and imitating complex articulatory vocalized sequences and exploiting their sonoric and rhythmical virtues is a separate capacity and has separate functions from the production of sound when governed by semantic uses as components of a lexicon. The theory of a prosodic protolanguage, thus, supposes that what emerged first in evolutionary history was an elaborated capacity for articulated vocal production. Only later did these sequences begin to acquire semantic functions, eventually developing into spoken language. Since we do not see this capacity for anything like a “bare phonology” in any of the apes (as far as is known, the complex sound sequences produced by gibbons are genetically pre-determined), we must suppose that this capacity developed in the hominid line, after the split from the Last Common Ancestor.

What might have led to the evolution of this capacity? Darwin may be credited with proposing that, in the light of the many parallels between human speech (as “bare phonology”) and song in song birds - parallels we alluded to earlier in this essay - the selection processes that led to song in birds might very well be paralleled in humans. He proposed that it was sexual selection that brought about song elaboration in birds (this is now widely accepted and there is now much evidence to support the idea) and suggested that this would also have operated among humans.

There are at least two difficulties for a sexual selection theory for the origins of a musical protolanguage in humans, however. First of all, the human capacity for speech production, imitation and learning is the same in both sexes. Traits that develop as a result of sexual selection usually develop to a much greater extent in one sex. In northern song birds, it is typically the male that has the elaborate song. Second, traits developed through sexual selection usually become manifest with the onset of sexual maturity - as is the case with singing in birds - but in humans, of course, the capacity for speech begins very early indeed. Thus, while it is probable that sexual selection played a role in the development and elaboration of the human voice - the differences between male and female voices in humans would suggest this - we must also look for other possibilities.

One possibility that Fitch favours derives from the importance of the voice, especially a singy-songy voice, in establishing and maintaining and elaborating the adult-infant relationship. Mothers all over the world sing to their infants, and singing games, word play of various kinds, play a very important role in interaction with infants, as has been pointed out by Ellen Dissanyake (2000). It has been suggested by Dean Falk (2004) that because the human infant is so very dependent at birth, but because, since humans lack fur, it cannot easily remain clinging to its mother for long periods (as chimpanzee infants can and do) but must be carried, for the mother to become free enough to engage in gathering food and other activities, it would be useful for the mother to “park” its infant. But to do so, would require that there be some way for close contact to be maintained. Falk suggests that this could have been an important factor favouring the emergence of complex, voluntary vocalisation, as part of a system of maintaining mother-infant contact. This idea not only solves the question of sexual equality in human vocal capacities (human males also look after their children and, in any case, adult-infant interaction is highly interactive, involving the baby as much as the adult), it also is consonant with the fact of very early development

of speech capacities in humans.

In addition to the adult-infant interaction hypothesis, there are a number of other considerations that could also be brought to bear on this issue. Fitch only mentions this in passing, but it seems to me he could have made more of the fact that there are many species of primates - such as certain species of baboon, and also some of the forest dwelling South American species, such as tamarins, who engage in extensive reciprocal chatter, chorusing and grunting, which serves an important role in maintaining contact between individuals in the group. This phatic use of reciprocated vocalisation is obviously of great importance and a comparative study of this kind of vocal usage in relation to the ecology and group structures of the species involved could suggest some further important hypotheses as to why complex vocal communication can be favoured in evolution, hypotheses that could very well apply to humans as well. Indeed, a recent study by McComb and Semple (2005), in which group size and complexity of social relations in the group were compared across several primate species, demonstrated an increase in the complexity of the vocal repertoires as group size and complexity of social structure developed. Here it is appropriate to mention again Robin Dunbar's "grooming" hypothesis for the origin of language (Dunbar 1996). As Fitch points out in his discussion of it (in Chapter 12), although Dunbar's treatment seems to take for granted the presence of propositional communicative functions and does not account for the origin of this aspect of language, his idea that vocal exchanges could substitute for grooming exchanges when group size extends beyond a certain limit, is a valuable additional hypothesis to account for the origin of musical or prosodic protolanguage, even if it is not useful for accounting for the origin of the semantic functions of modern language.<sup>28</sup>

Taking all this into consideration, there seems to be good reason to suppose that 'speech' (in the "bare phonology" sense of it, or "lallation" as I suggest we call it) had a separate origin in the hominid line, and comparative considerations, some of which we have outlined above, can suggest what might have been the selective pressures that favoured its emergence.

But how did this musical protolanguage come to have *meaning* in the sense that we understand that language today has meaning. How did speech become a vehicle for language? One route to a more language-like meaning that Fitch favours was originally suggested by Otto Jespersen (1922), whose contribution he describes in some detail. According to Jespersen, units of song-like utterance could have first acquired signification by being repeated in certain situations or by specific individuals, and could, thus, have become *leitmotifs* for these. Then, as they were repeated, they could, by being applied to associated circumstances or individuals, come to have a more general significance. For Jespersen, these meanings are *holistic* - units of utterance would acquire complex meanings as complete units, and would have functioned rather as very complex words do in some highly fusional languages, as in North America, where a single long word can have a meaning, which, if translated into English, would have to be rendered as an entire sentence.

In support of the idea that units in a protolanguage, as they acquired semantic

---

<sup>28</sup> See also the highly suggestive observations and discussions by Bruce Richman, who has made extensive studies of the vocal behaviour of gelada baboons (see Richman 1978, 1980, 1987)

significance, would have been holistic, Fitch appeals to the work of Allison Wray (1998, 2002). Wray does not discuss the relationship between music and language and does not entertain the idea of a musical protolanguage, but she argues that words were complex wholes originally and have become composed of separable parts, detaching nouns from verbs, developing particles and pronouns much later on. She points out that much spoken language has a holistic character today. We speak often in formulae, patterns of word sequences are often highly repeatable, as in social ritual interactions of all kinds. Furthermore, she points out, children, when acquiring language, often understand utterances as wholes, rather than as being divided into separate words. An objection that has been raised against this idea is that it might be difficult to envisage how, from the holistic complexes of a protolanguage, units would come to be separated and come to function as individual words. This issue has been raised with respect to how a child, encountering a parental language of utterances with only holistic meanings, would arrive at individual words. It has been suggested that individual words would arise from a process of over-generalisation by the child. Children, in acquiring language, often do employ meaningful units such as words in over-extended ways. For example, a child might use a word by which he designates a 'dog' to designate also a 'cat', generalising its meaning to cover all four footed creatures. It has been suggested, also, that children might over-analyse holistic utterances and generalise the results. This could result in the separation of words. We know that lexical innovations are often created by children, and sometimes they enter into the family lexicon; and we have also seen, from the work on Nicaraguan sign language, the important role children can play in creating innovations in linguistic systems. The idea that children, in acquiring a linguistic system, may change it in a certain way is thus not an idea without support.

As a kind of proof that this sort of process is plausible, Fitch turns to the work of Simon Kirby (2000) who, in a series of ingenious computer simulation experiments, has demonstrated how a compositional, syntactically structured language will emerge, simply if you start with a population of individuals, each of whom possesses the capacity to emit utterances composed of sequences of symbols that map as wholes on to meaning complexes. These individuals can learn, simply by observation. The utterances that each produces initially are for meanings that they are instructed to express, selected at random from the repertoire of meanings supplied. From time to time, some individuals are withdrawn from the population and replaced by new ones, who know nothing. This ensures there is always a new generation entering the population, who must learn from the language already established in previous generations. It is found that over a series of runs, a stable system of compositional utterances will emerge, with the parts of these utterances slowly coming to have stable meanings. As this happens the utterances come to be organised according to a simple syntax.

This demonstration by Kirby is meant to show that, starting with a 'phonological' system of re-combineable elements, and given that conceptual meanings can be mapped on to them, a lexicon with a compositional morphology will arise automatically, provided you build in certain assumptions about how the individuals in the system learn from each other. It shows how, starting from a non-syntactic holistic protolanguage, a syntactically organised system of utterances using stable lexical forms can emerge. This is offered by Fitch as an

answer to the critics of Jespersen and Wray who had claimed that this was not possible.

### **Discussion**

Fitch begins his book with the parable of an elephant exhibited in a dark room. Each person who approached the elephant could only experience it by touching it and, accordingly, depending upon what part of the elephant he touched and how he did so, he would arrive at a different conception of what the elephant was like. In approaching the problem of the origin of language, Fitch suggests, “[a]ll of us are still exploring the elephant of language in the darkness, all of us with only a partial understanding, and each discipline will have its place in the richer description and understanding that all are seeking” (p. 2). He insists (pp. 4-5) that language “must be viewed as a composite system, made up of many partially separable components.” He suggests that it “is not a monolithic whole, and from a biological perspective may be better seen as a ‘bag of tricks’ pieced together via a process of evolutionary tinkering.” He eschews the idea that there is just one aspect of language that is “core” or “central”. Nonetheless, although he says that many of the components that go to make up language are found in other animals (and this is what justifies the emphatically comparative approach he follows throughout the book) there are a few “core aspects of the human language capacity that remain unique to our species.” (p. 6).

Not surprisingly, it is these few “core aspects” that are the main focus of attention in the various models of language evolution that are evaluated in the last three chapters of the book - the various lexical protolanguage models, the gestural protolanguage models and the musical protolanguage model. As I have indicated in my discussion of these chapters, Fitch evaluates these models, and he points out the shortcomings of each of them. He says, however, that all of them also bring us valid insights and he suggests that we should not see them as conflicting. He is right about this, at least insofar as each model might be said to tackle the problem of language origins at different stages of development. The lexical protolanguage models already assume that humans had at their disposal units of expression that could serve as “words” in some sense. The gestural protolanguage models perhaps try to address the issue of how such “words” could have come about in the first place - Fitch suggests (p. 509) that these gestural models provide “the signalling prerequisites for a later ‘lexical protolanguage’” - while the musical protolanguage idea is an attempt to address the problem of providing an adaptive explanation for the discontinuity among primates in vocal learning. It also seeks to account for the relationships between musical and spoken uses of the voice and offers an approach which can account for the origin and importance of intonation tunes which serve such a fundamental role in speaking.

The emphasis of Fitch’s book, it should be said, is very much upon those aspects or capacities that individuals must have if they are to be able to ‘have’ language or to ‘do’ language (as it might be preferable to say). Fitch makes this explicit in his discussion of Chomsky’s distinction between “E-language” and “I-language”. According to this, ‘I-language’ is the neural cognitive system which exists within the individual that makes possible the ability to use language. E-language, on the other hand, as Chomsky saw it, is simply an aggregate epiphenomenon, no more than the output of a set of I-languages. Chomsky argued that I-language should be the proper object of study in the study of language. Fitch agrees with Chomsky “that

scientists interested in the genetic and neural mechanisms underlying language need to focus on I-language, as instantiated in individuals' brains." (p. 34). For him, thus, "I-language is the proper empirical starting point for this investigation" (*ibid*) - by which he means an investigation into the evolutionary origins of the various capacities and abilities humans have so that they can engage in languaging. However, Fitch agrees that E-language must also be investigated, and here he differs from Chomsky, for he does not dismiss E-language as an epiphenomenon. He says that the shared social systems that correspond to specific languages, such as English or French or Warlpiri are systems that become part of the human environment, they are part of the 'niche' that humans construct for themselves and they must, accordingly, have consequences for biological processes of human adaptation (p. 34).<sup>29</sup> This means that such systems are also properly within the purview of language origins studies. He recognises, however, that the processes by which these systems come about are *social* processes, processes that he refers to as *glottogenic*, and these are not fully accountable in biological terms.

Since it is the biological processes that are the principal emphasis of the book, these social processes receive rather brief discussion. For example, as we have seen, Fitch recognises that for a fully developed theory of protolanguage, as discussed in Chapter 12, a 'two-stage' theory is necessary. A second stage is needed to account for the fact that the propositional information that utterances can convey, can be shared beyond circles of immediate kin. Fitch also recognises, as we see in Chapter 3, where he outlines the main features of 'language', that in order to understand how linguistic utterances can have meaning, they must be understood in terms of the inferences they allow their users to make about what each had intended to convey. This, of course, has implications for the kinds of cognitive abilities that are required for being able to 'do language'. In developing this theme, he devotes some space to Gricean theory and to Sperber and Wilson's elaboration of this.<sup>30</sup> He relies on Grice to bring out the point that, in human linguistic communication, participants *cooperate* with one another in the light of "an essential common interest in getting some point across" (p. 134). He suggests that this common desire of participants to communicate cooperatively is peculiar to human communication and is not found anywhere else in the animal kingdom. What demands explanation, from an evolutionary point of view, according to Fitch, is how it is that human speakers are able to modulate their communicative signals according to the information that they know that this will provide their conspecifics and that they can do this in the light of their understanding of what information it is that their co-participants need so that they can collaborate in the realisation of a jointly entertained goal.

This issue of how participants in interaction can come to share a common perspective is central to understanding how shared referentiality is possible. Fitch's approach to this question is decidedly 'cognitivist'. He accepts the idea that humans develop what has come to be called a "theory of mind" and it is in terms of this that each other's behaviour is interpreted. Although widely followed, this approach depends upon a number of assumptions that may be open to question, and it does not direct attention to the detailed

---

<sup>29</sup> Bickerton (2009) makes a similar point.

<sup>30</sup> Sperber and Wilson (1995).

study of how humans actually accomplish interaction when they are co-present.<sup>31</sup> A considerable body of work has now been carried out in which the way in which ongoing interaction is actually organized has been described and from this it seems clear that an account in terms of the kind of refined inferential processes that a “theory of mind” approach implies is not necessary. Behaviour is highly patterned and humans (and not only humans) are immersed in this patternedness from the beginning.

We can gain some clues regarding this from the work of Erving Goffman (e.g. Goffman 1961, 1963, 1973) whose work is not usually mentioned in the context of language origins discussions. From this we understand how people characteristically enter together into occasions of what he has called “focused interaction” in which they jointly agree upon what is relevant for the occasion and what may be disattended. Such jointly sustained attentional frames seem to be a fundamental feature of coherent interactions of any kind and it is only by seeing how communicative exchanges depend upon the creation of such jointly constructed shared “micro-worlds” that we can come to see how mutual understanding is achieved.

Approaching this problem from a somewhat different perspective, this point is similar to the one that Tomasello (2008) has been arguing for: that shared referential understanding can only come about within a joint attentional framework. That is, for an action to have common referential significance it is necessary for the participants to somehow share an understanding that they are both attending to the same things in the same way.

The achievement of this joint attentional frame may be accomplished in a variety of ways, but in fully co-present or non-mediated interactions much depends upon delicate coordinations between movements and orientations of the participants. It is through such coordinations that shared cooperative intentions can become manifest for the participants and so be established among them. In some of my own work of some years ago (Kendon 1985, reprinted in Kendon 1990: 239-262), for example, I described how the spatial-orientational systems that participants in focused interaction can enter into and cooperatively sustain, play an important role in the means by which is achieved the attentional “frame attunement” necessary for the common understanding in terms of which participants’ actions make sense. This need not be done by words or by gestures, but by reciprocally sustained spatial and orientational manoeuvres. Accordingly, when it is seen that intelligible linguistic exchanges pre-suppose and depend upon the setting up of such joint attunements, we come to see that the very activity of uttering linguistic acts of some sort can only be understood when the setting up of interactional settings, the establishment of “participation frameworks” is also understood. There are now a number of good descriptions of this for human interaction. Studies of great ape interaction that take a comparable approach would be extremely useful. A few beginning steps have been made, for example in the work of Simone Pika (Pika and Mitani 2009) with chimpanzees, the work of Joanne Tanner on gorillas (Tanner 2004) and see also the book by Barbara King *The Dynamic Dance* (King 2004). Such work will allow us to compare the organisation of occasions of interaction between species, not just the vocal and gestural signals they produce as discrete units of action (a common approach hitherto, see Call and Tomasello 2007), and this will

---

<sup>31</sup> For critiques of Theory of Mind see, for example, Leudar, Costall and Francis (2004) and the papers that follow in the same Special Issue of *Theory and Psychology*

greatly enrich our understanding of the circumstances in which the emergence of joint referentiality of actions, in whatever modality, might have been enabled. Almost nowhere in the language origins literature are issues of this sort discussed.

Another important feature of languaging, already alluded to above, is the fact that when speakers construct utterances they *always* do so through an orchestration of diverse semiotic resources (Goodwin 2000). Now although acknowledgement is often given to the fact that in speaking speakers also make use of “paralanguage” - intonational modulations in speech production and various kinds of kinesic accompaniments, these generally tend to be treated as *auxiliary or decorative accompaniments* and not as integral to the very activity by which an utterance is produced. For example, Fitch says (p.509) that modern humans in all cultures “use gestures meaningfully” and adds, accordingly (noting that apes also use gestures “meaningfully”)<sup>32</sup>, that “we have every reason to believe ... that gesture played an important supporting role in communication throughout hominid evolution.” But what is meant here by “supporting”? Should we not be impressed by the fact that verbal language, when produced by speakers, is *never* produced as only words?<sup>33</sup> We can, of course, write down a person’s words, presenting their utterances so that they seem to be made up only of words. However this does not represent what was actually done when the utterance was produced. Whenever a person speaks he employs, in a completely integrated fashion, patterns of voicing and intonation, pausings and rhythmicities, which are manifested not only audibly, but kinesically as well. Always there are movements of the eyes, the eyelids, the eyebrows, the brows, as well as the mouth, and patterns of action by the head. In addition there are, from time to time, variously conspicuous hand and forearm actions or ‘gestures’ (as they are often called), as well as postural and orientational changes. All of this is produced in full orchestration with speech - complex and variable, to be sure, but always orchestrated - and must be seen as inseparable components of the utterance as the utterer produces it.<sup>34</sup> Few theorists of language offer an account of this.<sup>35</sup> This may be, of course, because most theorists of language hitherto have not seen it as their business to do so since ‘language’ in most such cases is thought of as a self-contained, autonomous system that is confined to only one modality. But is this view of language anything other than a convenient abstraction? And if so, does it then constitute an appropriate target for evolutionary explanations?

This question, posed a long time ago by Bolinger in his “Thoughts on ‘yep’ and ‘nope’” published in 1946 (Bolinger 1946), has been posed again more strikingly, perhaps, as

---

<sup>32</sup> Fitch seems to overlook the fact that the way humans use gestures “meaningfully” is quite different from the way apes use them. Apes probably do not point and they have never been seen to use their hands to describe things, which is a very common kind of gesture use in humans!

<sup>33</sup> Some of what follows here has been adapted from my essay “Language’s matrix” (Kendon 2009)

<sup>34</sup> Early demonstrations of this were by Condon and Ogston (1967), see also various chapters in Birdwhistell (1970) and Kendon (1972). Detailed discussions of gesture in relation to speech are found in McNeill (1992) and Kendon (2004).

<sup>35</sup> Fitch does make reference to Donald (1991), who discusses speech co-ordinated gesture, which he agrees has undergone evolutionary change along with speech. He links it to an earlier mimetic stage in the evolution that was to lead to language, and Fitch (p. 505) favours the idea of a “multimodal communication stage”. Putting it this way, however, leaves the impression that multimodality has been or is being left behind. This is far from the case, and the relevance of this for theories of language evolution has yet to be adequately developed.



a result of the recent work on sign languages (Liddell 2003: 355-362). When, after William Stokoe's demonstration in 1960 that the "visual communication system of the deaf" (as he called it) could be analysed, at least to a considerable extent, in terms of the analytic principles developed for spoken languages by Bloomfield and his followers, such as George Trager (who directly taught Stokoe), there developed a determination to demonstrate that such structuralist principles were completely adequate for the analysis of sign languages, for in this way it would be shown that sign languages were indeed *languages*, and not, as had been maintained for the prior eighty years or so, "mere pantomime" or "unsystematic gesture". In doing this, the concept of 'language' as a self-contained system was extended to include sign languages which meant that they also came to be conceived of as well demarcated autonomous systems.

However, because there is no tradition of writing for sign languages and so no pre-established criteria for deciding what is "in" the language and what is "outside" it, any attempt to suppose that there can or should be a separation between 'paralinguistic' features and 'linguistic' features becomes very problematic. In recent years it has become clear that central to the construction of utterances in sign languages are forms of expression such as 'classifiers', 'constructed action', or "highly iconic forms" (on this, see Cuxac and Sallandre 2007), as well as an exploitation of space that is not possible in speech, but which have much in common with various kinesic devices used by speakers (although these are not as systematic in speakers).<sup>36</sup> This proves to be an embarrassment to those who want to maintain a model of sign language that is derived from existing models in spoken language linguistics. On the other hand, this has also led others to suggest that when comparing spoken and signed language, the comparison should be with language as it is performed in speaking, for it is only the performed version of a sign language that is ever available. If this suggestion is followed, however, this means that, after all, from the point of view of how utterances are constructed, it is as essential to view what speakers do as an integrated performance in the study of spoken languages, as it is in the study of sign languages.

In short, we may suggest that the 'natural' state of spoken language is a *speech-kinesis ensemble*. Presumably, this has always been the case. With the development of writing, and its ultimate emergence as an autonomous form of language with its own properties, which, nevertheless, has provided the dominant model for what 'language' is, at least since the end of the eighteenth century, we have ceased to see how gestures and other aspects of utterance performance are a part of "what is said." In (relatively) recent history, as human cultures have developed to sustain ever larger units of social organisation, especially, we repeat, with the development of writing technologies, the separation and specialisation of modalities of communication has been favoured. In many glottogenetic discussions, it is the separated modality of written-down spoken language (which dominates our conception of language) that tends to be projected backwards to the earliest days of language, making it very difficult to imagine how it might have arisen.

We may suppose, however, that just as 'linguaging' is, in fact, a poly-modal activity today, so it must have been in its beginnings (incidentally, to take this point of view

---

<sup>36</sup> See Kendon (2004: Ch. 15).

obviates the problem that “gesture first” scenarios have raised)<sup>37</sup>. This leaves us with the question as to why there is this poly-modality and why, in particular, when speakers speak (or signers sign, for that matter) they tend to mobilise all kinds of bodily resources beyond those that might seem necessary from a mono-modal point of view. The model of ‘language’ as an autonomous mono-modal system, which tends to be taken as the target in so many language origins discussions, is a system that is a product of latter-day reflections on language, greatly facilitated ever since systems of writing came to be regarded mainly as representations of spoken language. A model of language of this kind is not appropriate to apply in those primordial times when what were to become specifically human forms of communication were first emerging. ‘Language’, as it is so often conceived of in contemporary language evolution discussions, is a late differentiation from a complex and dynamic orchestration of communicative action. Furthermore, it is a continually emerging system. The ‘target’ of our evolutionary explanations perhaps should be re-formulated so that the poly-modality of utterance is taken as the starting point. If this is done then ‘language’, when it is considered in its mono-modal form, can then be understood as a consequence of processes of specialisation and differentiation from poly-modal ensembles of action. Accounts of language origins can then be recast to become accounts of these processes of progressive specialisation and diversification, emerged and emerging systems that are shaped through an evolution that involves social interaction as much as biology.

## References

- Arbib, M. (2005). From monkey-like action to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105-167.
- Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the Nature of Language*. Cambridge: Cambridge University Press.
- Bellugi, U. (1981). The acquisition of a spatial language. In F. S. Kessel (Ed.), *The Development of Language and Language Researchers: Essays in Honor of Roger Brown* (pp. 153–185). Hillsdale, New Jersey: Lawrence Erlbaum.
- Bickerton, D. (1981). *The Roots of Language*. Ann Arbor: Karoma Publishers, Inc.
- Bickerton, D. (1990). *Language and Species*. Chicago: Chicago University Press.
- Bickerton, D. (1995). *Language and Human Behavior*. Seattle, WA: University of Washington Press.
- Bickerton, D. (2009). *Adam's Tongue: How Humans Made Language, How Language Made Humans*. New York: Hill and Wang, A Division of Farrar, Strauss and Giroux.
- Birdwhistell, R. L. (1970). *Kinesics and Context: Essays in Body Motion Communication*. Philadelphia: University of Pennsylvania Press.
- Bolinger, D. (1946). Some thoughts on 'yep' and 'nope'. *American Speech*, 21, 90-95.
- Botha, R. & Knight, C. (Eds.). (2009). *The Cradle of Language*. Oxford: Oxford University Press.
- Botha, R. & Knight, C. (Eds.). (2009). *The Prehistory of Language*. Oxford: Oxford University Press.

---

<sup>37</sup> See also Kendon (2011).

- Bronowski, J. & Bellugi, U. (1970). Language, Name, and Concept. *Science*, 168, 669–673.
- Brown, S. (2000). The 'musilanguage' model of language evolution. In N. L. Wallin, B. Merker, & S. Brown (pp. 271-300). Cambridge, MA: MIT Press.
- Call, J. & Tomasello, M. (Eds.) (2007). *The Gestural Communication of Apes and Monkeys*. Mahwah, NJ: Lawrence Erlbaum, Associates.
- Carpenter, C. R. (1940). A field study in Siam of the behavior and social relations of the gibbon (*Hylobates lar*). *Comparative Psychology Monographs*, 16 (5).
- Catchpole, C. K. & Slater, P. J. (2008). *Bird Song: Biological Themes and Variations*. Second Edition. Cambridge: Cambridge University Press.
- Cheney, D. L. & Seyfarth, R. M. (1990). *How Monkeys See the World*. Chicago: Chicago University Press.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton and Co.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: Massachusetts: MIT Press.
- Condon, W. S. & Ogston, W. D. (1967). A segmentation of behavior. *Journal of Psychiatric Research*, 5, 221-235.
- Corballis, M. C. (2002). *From hand to mouth : the origins of language*. Princeton: Princeton University Press.
- Cuxac, C. & Sallandre, M.-A. (2007). Iconicity and arbitrariness in French Sign Language: Highly iconic structures, degenerated iconicity and diagrammatic iconicity. In E. Pizzuto, P. Pietandrea, & R. Simone (Eds.), *Verbal and Signed Languages: Comparing Structures, Concepts and Methodologies* (pp. 13-33). Berlin: Mouton de Gruyter.
- Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
- Dissanayake, E. (2000). Antecedents in the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 389-410). Cambridge, MA: MIT Press.
- Falk, D. (2004). Prelinguistic evolution in early hominims: Whence motherese? *Behavioral and Brain Sciences*, 27, 291-541.
- Gardner, R. A. & Gardner, B. (1969). Teaching sign language to a Chimpanzee. *Science*, 165, 664–672.
- Gentilucci, M. & Dalla Volta, R. (2007). The motor system and the relationship between speech and gesture. *Gesture*, 7 (2), 159-177.
- Goffman, E. (1963). *Behavior in public places*. New York: The Free Press of Glencoe.
- Goffman, E. (1961). *Encounters*. Indiannapolis: Bobbs-Merrill.
- Goffman, E. (1963). *Behavior in Public Places. Notes on the Social Organization of Gatherings*. New York: Free Press of Glencoe.
- Goffman, E. (1974). *Frame analysis*. Cambridge, MA: Harvard University Press.
- Goffman, E. (1981). *Forms of talk*. Philadelphia: University of Pennsylvania Press.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, 1489-1522.
- Harnard, S., Steklis, H. D., & Lancaster, J. (Eds.). (1976). *Origin and evolution of language and speech*. New York: New York Academy of Sciences.
- Hauser, M. D. *The Evolution of Communication*. Cambridge, MA: MIT Press.
- Hayes, C. (1951). *The Ape in Our House*. New York: Harper.

- Hewes, G., Stokoe, W. C., & Wescott, R. W. (Eds.). (1974). *Language Origins*. Silver Spring, MD: Linstok Press.
- Hewes, G. W. (1973). Primate communication and the gestural origins of language. *Current Anthropology*, 14, 5-24.
- Hockett, C. (1960). Logical considerations in the study of animal communication. In W. E. Lanyon & W. N. Tavolga (Eds.), *Animal Sounds and Communication*. Publication No. 7 of the American Institute of Biological Sciences (pp. 392–430). Washington, D. C.: American Institute of Biological Sciences.
- Hockett, C. & Ascher, R. (1964). The human revolution. *Current Anthropology*, 5, 135–168.
- Hurford, J. R., Studdert-Kennedy, M., & Knight, C. (Eds.). (1998). *Approaches to the Evolution of Language*. Cambridge: Cambridge University Press.
- Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, 3 (7), 272-279.
- Jespersen, O. (1922). *Language: Its nature, development and origin*. New York: Henry Holt and Co.
- Johansson, S. (2005). *Origins of Language: Constraints on Hypotheses*. Amsterdam and Philadelphia: John Benjamins.
- Jurgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Review*, 26, 235-258.
- Jurgens, U. (2000). A comparison of the neural systems underlying speech and nonspeech vocal utterances. In B. H. Bichakjian, T. Chermigovskaya, A. Kendon, & A. Moller (Eds.), *Becoming Loquens: More Studies in Language Origins* (pp. 1-13). Frankfurt am Main: Peter Lang.
- Kegl, J., Senghas, A., & Coppola, M. (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. DeGraff (Ed.), *Language Creation and Language Change. Creolization, Diachrony and Development* (pp. 179–237). Cambridge, Mass.: MIT Press.
- Kellogg, W. N. & Kellogg, L. A. (1933). *The Ape and the Child*. New York: McGraw Hill.
- Kendon, A. (1972). Some relationships between body motion and speech. An analysis of an example. In A. Siegman & B. Pope (Eds.), *Studies in Dyadic Communication* (pp. 177-210). Elmsford, New York: Pergamon Press.
- Kendon, A. (1990). *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press.
- Kendon, A. (2002). Historical observations on the relationship between research on sign languages and language origins theory. In D. Armstrong, M. A. Karchmar, & J. V. Van Cleve (Eds.), *The Study of Signed Languages: Essays in Honor of William C. Stokoe*. (pp. 32-52). Washington, D. C.: Gallaudet University Press.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kendon, A. (2008). Signs for language origins? *Public Journal of Semiotics*, II (2): 2-29.
- Kendon, A. (2009). Language's matrix. *Gesture*, 9 (3), 352-372.
- Kendon, A. (2011). 'Gesture first' or 'Speech first' in language origins? In D. J. Napoli & G. Mathur (Eds.), *Deaf Around the World* (pp. 25-267). New York: Oxford University Press.
- King, B. J. (2004). *The Dynamic Dance*. Cambridge, MA: Harvard University Press.

- Kirby, S. (2000). Syntax without natural selection. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Functions and the Origins of Linguistic Forms* (pp. 303-323). Cambridge: Cambridge University Press.
- Klima, E. A. & Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.
- Knight, C., Studdert-Kennedy, M., & Hurford, J. R. (Eds.). (2000). *Evolutionary Emergence of Language*. Cambridge: Cambridge University Press.
- Lenneberg, E. (1967). *Biological Foundations of Language*. New York and London: John Wiley and Sons.
- Leudar, I., Costall, A. & Francis, D. (2004). Theory of Mind: A Critical Assessment. *Theory and Psychology*, 14 (5), 571-578.
- Liddell, S. K. (2003). *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
- MacNeilage, P. F. (2008). *Origin of Speech*. Oxford: Oxford University Press.
- McComb, K. & Semple, S. (2005). Coevolution of vocal communication and sociality in primate. *Biology Letters*, 1 (4), 381-385.
- McNeill, D. (1992). *Hand and Mind: What Gesture Reveals about Thought*. Chicago: Chicago University Press.
- Mithen, S. (2009). Holistic communication and the co-evolution of language and music: resurrecting an old idea. In R. Botha & C. Knight (Eds.), *The Prehistory of Language* (pp. 58-76). Oxford: Oxford University Press.
- Odling-Smee, J. & Laland, K. N. (2009). Cultural niche construction: evolution's cradle of language. In R. Botha & C. Knight (Eds.), *The Prehistory of Language* (pp. 99-121). Oxford: Oxford University Press.
- Paget, R. A. S. (1930). *Human Speech. Some Observations, Experiments, and Conclusions as to Nature, Origin, Purpose and Possible Improvement of Human Speech* (AMS Reprint 1978 ed.). New York: Harcourt, Brace and Company.
- Pika, S. & Mitani, J. C. (2009). The directed scratch: evidence for a referential gesture in chimpanzees? In R. Botha & C. Knight (Eds.), *The Prehistory of Language* (pp. 166-180). Oxford: Oxford University Press.
- Pike, K. L. (1967). *Language in Relation to a Unified Theory of the Structure of Human Behavior*. The Hague: Mouton.
- Pinker, S. & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707-784.
- Radick, G. (2008). *The Simian Tongue*. Chicago: University of Chicago Press.
- Richman, B. (1978). The synchronization of voices by gelada monkeys. *Primates*, 19, 569-581.
- Richman, B. (1980). Did human speech originate in coordinated vocal music? *Semiotica*, 32, 233-244.
- Richman, B. (1987). Rhythm and melody in gelada vocal exchanges. *Primates*, 28, 199-223.
- Rizzolatti, G. & Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188-194.
- Rizzolatti, G. & Sinigaglia, C. (2006). *So quel che fai: il cervello che agisce e i neuroni specchio*. Milan: Raffaello Cortina Editore.

- Sche en, A. E. (1972). *Communicational structure: Analysis of a psychotherapy transaction*. Bloomington: Indiana University Press.
- Sebeok, T. A. & Umiker-Sebeok, D.-J. (Eds.). (1980). *Speaking of Apes*. New York and London: Plenum Press.
- Sperber, D. & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell Publishing.
- Stam, J. (1976). *Inquiries into the Origin of Language: The Fate of a Question*. New York and London: Harper and Row.
- Stokoe, W. C. (1960). Sign Language Structure: An outline of the visual communication systems of the American deaf. *Studies in Linguistics Occasional Papers*, 8, 1-78.
- Stokoe, W. C. (2001). *Language in Hand*. Washington, D. C.: Gallaudet University Press.
- Tanner, J. E. (2004). Gestural phrases and gestural exchanges by a pair of zoo-living lowland gorillas. *Gesture*, 4 (1), 1-24.
- Tomasello, M. (2008). *The Origins of Human Communication*. Cambridge, MA: MIT Press.
- Wallman, J. (1992). *Aping Language*. Cambridge: Cambridge University Press.
- Woll, B. (2009). Do mouths sign? Do hands speak?: Echo phonology as a window on language genesis. In R. Botha & H. de Swart (Eds.), *Language Evolution: The View from Restricted Linguistic Systems* (pp. 203-224). Utrecht: LOT.
- Wray, A. (1998). Protolanguage as a holisitic system for social interaction. *Language and Communication*, 18 (1), 47-67.
- Wray, A. (2002). Holistic utterances in protolanguage: the link from primates to humans. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The Evolutionary Emergence of Language* (pp. 285-302). Cambridge: Cambridge University Press.