

# LINGUISTICS FROM AN EVOLUTIONARY POINT OF VIEW

James Hurford

## 1 LINGUISTICS AND EVOLUTION

Beginning Linguistics students are sometimes treated to an array of mock “theories” about the evolution of language, including the “Bow-wow” theory, the “Ding-dong” theory and others with equally silly and dismissive names. The 1886 ban on the subject (along with proposals for a universal language) by the Société Linguistique de Paris is well known, and also sometimes thrown up as a proscription that should be reimposed. Research into the evolution of language never really died, though its serious contributors, such as C.F. Hockett [1960] and Philip Lieberman [1984], were tiny in number. In the past twenty years, the resurrection of the subject has accelerated dramatically. The resurgence can be attributed to a general increase in multidisciplinary research, and to impressive empirical advances in relevant fields such as genetics, psychology of language, ethology (especially primatology), computer modelling, linguistics (especially language typology and some formal modelling) and neuroscience.

Linguistics has traditionally been isolated from evolutionary considerations. Saussure’s [1916] emphasis on the primacy of synchronic descriptions coloured all of mainstream 20<sup>th</sup> century Linguistics. The core of generative grammar is synchronic work. Moreover, the emphasis in generative theory on the discovery of abstract formal principles governing the shape a language can take tends to isolate the study of language from neighbouring disciplines. The prevailing assumption within this dominant paradigm has been that the principles to be discovered are peculiar to language alone [Chomsky, 1965; 1975; 1981]. If regularities are observed that can be accounted for in terms of more general human behaviour, or even animal behaviour, such as memory limitations, limitations on the physiology of the output device (vocal or manual), or constraints on processing complexity, these have tended to be sidelined as not within the domain of Linguistics proper, which is taken to be whatever is special to language alone. There is more than a whiff of Platonism in much 20<sup>th</sup> century theorizing about language. Of course, as Linguistics is a large field, there have been dissenting voices (e.g. [Bybee, 1985; 1994; Givón, 1979; 1990]), emphasizing the integration of the study of language structure with the study of human and animal behaviour generally, and taking a

more favourable attitude to explanations in terms of function (as opposed to an appeal to deep-seated abstract principles not necessarily motivated by function).

Historical linguists, though working with diachrony, have almost universally taken a uniformitarian stance, postulating that reconstructed proto-forms of languages are no different in kind from modern languages. In a uniformitarian view of this kind, change in language cycles around through states of language that are all of the same recognizably modern type. This is consistent with the standard teaching in Linguistics that there are no primitive languages. Thus the idea of languages evolving from earlier types different from the types observed today does not get an airing. The exception is Creole studies, where it is often acknowledged that these newly emerging languages are in some senses simpler than languages with long histories.

The isolation of mainstream Linguistics from evolutionary considerations is puzzling in light of Chomsky's emphatic and influential re-location of Linguistics within Psychology and ultimately Biology. Human language is a product of human minds and bodies, and these in turn are products of evolution. Chomsky and his fellow-thinkers do not deny that the human language capacity has evolved; rather, the argument is that the course of this evolution has not been significantly affected by natural selection. Whatever it was that gave humans this impressive capacity, setting us off very distinctively from other species, cannot (the argument goes) be attributed to incremental pressure to mould a system well adapted to communication. These dominant views were challenged influentially in 1990 by Pinker and Bloom, under the eloquent title "Natural Language and Natural Selection". Pinker and Bloom took their cue from mainstream generative grammar, whose methodology they accepted, and in which the tenor of the day was still that humans are born with an innate richly structured cognitive subsystem accounting for the rich complex structures of languages so easily acquired against the odds by all non-pathological children. They likened the complex structure of a human language to the complex structure of the human eye. Both, they argued, are the products of natural selection, working gradually. The challenge was to find some selective rationale for each separate component of the assumed many-faceted innate language acquisition device.

The fifteen years following Pinker and Bloom's article witnessed a spectacular reversal of the central theme in generative theorizing. Rather than the human language faculty being innately rich in detailed structure, a 'Minimalist Program' emerged [Chomsky, 1995]. From the viewpoint of language evolution, the most important manifestation of this movement was seen in an article co-authored by Hauser, Chomsky and Fitch [2002]. In this article, they usefully distinguished between the human language faculty in a broad sense (FLB) and the human language faculty in a narrow sense (FLN). FLN, that which is distinctive of human language, when compared to animal communication and to non-linguistic cognition, may consist, at most, of a capacity for recursion. That's all, and maybe FLN is even empty. In this latter case (FLN is null), what makes humans capable of language may be just the capacity to apply recursion to their communication sys-

tems; animals may conceivably be able to do some recursive computation (maybe in their navigation), but they don't use it in their communication.

The suggestion that the human language faculty in the narrow sense (FLN) is minimal is attractive to biologists and evolutionary theorists because there is less to account for. We don't have to find special selective rationales for a whole set of apparently arbitrary principles of an assumed innate complex template for language structure ('universal grammar', UG), peculiar to humans. Nevertheless, it remains the case that human phenotypic communicative behaviour is vastly more complex than anything in the non-human animal world. Scientific methodology, in linguistics as in biology, dictates that we postulate as little in our explanatory story as necessary. Somehow, we have to find plausibly little evolutionary changes that generated, perhaps in a cascade of subsequent changes, the massive difference we see today. And if we view language broadly, addressing the human faculty of language in the broad sense (FLB), much of the evolutionary basis for language can be sought in animal behaviour and human non-linguistic cognition.

The two major contenders for crucial evolutionary changes leading to modern language-using humans have been (1) a capacity for complex syntax, and (2) a capacity to learn tens of thousands of arbitrary symbols. The assertion that human syntax is complex is impressionistic, since it is not backed up by any quantitative metric, but the impression is surely nevertheless correct. In section 6 below on syntax, the most complex examples of syntax in non-humans will be briefly described. With ingenuity, it may be possible to reduce the syntactic complexity of languages to the interaction of recursive operations with somewhat complex lexical structure. Both complex syntax and vocabulary are acquired impressively fast and despite a poverty of the stimulus. Children produce complex sentences that they have never heard before, and acquire new lexical items on the basis of as few as one exposure in context. A third contender for a crucial difference between humans and non-humans is social trust, a factor that I will mention further in section 3, on Pragmatics.

A modern child is born into a society with a rich historically developed language, and internalizes most of this historical product in less than a decade. The ability to do this is a biological genetic endowment, which must have evolved, though we don't know in detail how it happened or how long it took. Being a matter of biological evolution, it was relatively slow, possibly taking millions of years (how many millions depending on how far back you start counting). Contrasting with this slow biological evolution of the human language faculty is the historico-cultural evolution of particular languages. The very first communicative codes used by biologically modern humans were presumably extremely simple, without the elaborate structure we see in modern languages. The pre-historical evolution of languages in communities of biologically modern humans was subject to the same range of pressures as are considered today by historical linguists studying the much more recent history of languages. Analogously, the laws of physics acting in the formation of the Earth over four billion years ago were the same as the laws of physics acting today, but the states of the Earth then and now are very different.

The pressures on language include economy of effort on the part of the speaker, balanced by a need to get ones meaning across clearly and distinctly, learnability, and usefulness in the normal arena of use (which itself evolves). Child creators of modern Creole languages are in a privileged position compared to the earliest humans who began to use language. Our remote ancestors had no model to learn from. There had to be invention, in some informal sense of that term. The early stages in the evolution of modern languages by cultural learning over successive generations would have been very slow at the start, some time between 200,000 and 100,000 years ago. It probably speeded up exponentially over the centuries. Certainly the ancient classical languages we know of today, less than 10,000 years old, look completely modern. To summarize, there are two senses of “evolution of language”. One is the relatively slow biological evolution of humans up to a language-ready stage; the other is the historico-cultural evolution of particular languages. A possibility that now seems increasingly plausible is that there has been a certain amount of gene-language coevolution. Given some cultural development of shared symbolic communication systems, the use of which conferred advantage both on groups and individuals, genes would have been selected enabling more efficient use of such systems. Such biological adaptations to an incipient linguistic culture would have favoured faster processing and increased memory storage for mappings between forms and their meanings.

Research into the evolution of language is unlikely to provide answers to the most commonly asked, and most naïve, factual questions. The most common, and most naïve, at least from laypeople, is “What was the first language?”. Linguists are rightly dismissive of this naïve question, as the techniques of historical reconstruction lose their power after at most 10,000 years. A minority of researchers (e.g. [Ruhlen, 1994]) claim to be able to reconstruct at least a few lexical items of “Proto-World”, the putative mother language of all modern languages. This work is very widely rejected by linguists, especially historical linguists, who argue that the statistical effects of merely coincidental change are not properly considered, and that the sampling methods on which the reconstructions are based are unsound. Less naïve questions, such as, for example, “Did *Homo neanderthalensis* have a language capacity comparable to *Homo sapiens*?”, or “At what stage in human pre-history did subordinate clauses first appear?”, are not now answerable, and may never be answerable.

So what kinds of questions do researchers in the evolution of language address? The key, I believe, is to take a cue from the dictum of the evolutionary biologist Dobzhansky [1973], who wrote “Nothing in biology makes sense except in the light of evolution”. In parallel, I claim, nothing about language makes sense except in the light of evolution. Linguists, for the most part firmly embedded in a synchronic paradigm, tell us in great detail what individual languages are like, and generalize, with the help of developmental psychologists, to what the innate human language faculty is like. These are descriptions of the current historico-cultural and biological states of affairs. Evolutionary linguistics adds an explanatory dimension. Of both kinds of state of affairs, the biological and the historico-cultural, we pose

the additional question, “And how did things get to be that way?”. Chomsky can be credited with adding the dimension of explanatory goals, as opposed to merely descriptive goals, to linguistics. The Chomskyan type of answer to “Why are languages the way they are?” is an appeal to innate dispositions in the language-learning child. This presupposes an existing complex language, to which the child is exposed. At the historico-cultural level, evolutionary linguistics adds the question “And how did existing complex languages get to be the way they are?”. At the biological level, the relevant additional question is “And how did the human species get to be the language-ready way it is?”. Obviously, these are contentful (if difficult) questions; they lead us away from the inward-looking study of language on its own, to a wider perspective of language in the context of perception, cognition and social arrangements, in non-human animals as well as in humans. The inward-looking theories of more conventional linguists must eventually be compatible with the wider perspective afforded by taking evolution, both biological and cultural, into account.

A language, conceived broadly (i.e. taking an FLB viewpoint) is a bridge between meanings and sounds (or manual gestures), and the meanings and sounds are themselves parts of the system, the end supports of the bridge, to pursue the metaphor. It will be convenient here to discuss this bridging system in terms of the usual compartments posited in linguistic theory, namely pragmatics, semantics, syntax, phonology and phonetics, as shown in Figure 1.

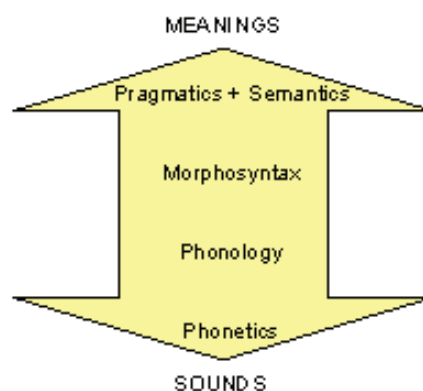


Figure 1. A language system is a bi-directional bridge between meanings and sounds. Linguistics naturally carves this bridge into the structurally different components identified here.

Viewing linguistic structure from an evolutionary point of view, one asks of each separate structural component of a language system “How did it get to be that way?” This leads one to consider possible evolutionary antecedents to the subsystem in question, which in turn leads to the recognition that, from an evo-

lutionary point to view, some modification of the traditionally defined boundaries between parts of a language system is appropriate. Thus the separate sections that follow will each begin with a definition of the relevant domain (e.g. pragmatics, phonetics) which is convenient from an evolutionary perspective.

The various subsystems of an overall language system are of varying antiquity and provenance. Some aspects of human linguistic behaviour are extremely ancient, shared with many mammals as part of a common biological heritage. Other aspects are also biologically determined, but special to humans, and are therefore more recent evolutionary developments. Finally, much human linguistic behaviour is learned, and culturally transmitted across generations; such parts of a language have evolved historico-culturally. For some conceptual clarity, in Figure 2 I give an extremely schematic view of the varying antiquity and provenance of subcomponents of a language system. This diagram should be taken with great caution; it is meant to be suggestive only, and there is no way in which such a diagram can be properly scaled. In what common quantal units, for instance, can one 'measure' the relative 'sizes' of the semantic and phonetic components of a language? It is clearly not possible. Yet I hope that this diagram will at least serve as a mnemonic for a general message, namely that the more peripheral parts of a language system, those dealing directly with meanings (pragmatics and semantics) and with sounds (phonetics) are the more ancient, and more rooted in our common biological heritage with other mammals. The inner parts of a language system, those having more to do with the formal organization and distribution of morphemes, words, phrases, clauses and sentences (i.e. morphosyntax) and of syllables and phonetic segments (i.e. phonology) have substantial learned components; thus the body of the syntax and phonology of a language, which must be learned by a child, has evolved historico-culturally, though it is still enabled by specific relevant biological capacities. As languages have grown culturally, there has also been a degree of biological co-evolution, adapting the organism to cope with the new complexities. In Figure 2, these new culturally-driven biological adaptations are represented by blocks labelled 'NEW BIO'.

## 2 SEMANTICS FROM AN EVOLUTIONARY POINT OF VIEW

Semantics is usually, within Linguistics, distinguished from pragmatics by not involving language-users in communicative situations. Traditionally, semantics has been defined as involving the truth-conditions of sentences and the denotations of words and phrases, considered out of context. This does not mean that linguistics ignores communication between people; this is dealt with under the heading of pragmatics. From an evolutionary point of view it is useful to keep half of the traditional core of semantics, namely the notion of a relation to entities in a world outside language, populated by objects, their static and dynamic properties, relations between objects, events involving such objects, and so on. Thus semantics is viewed as essentially extensional, involving a relation to an outside world. But if we are interested in the evolution of language as a psychobiological phenomenon,

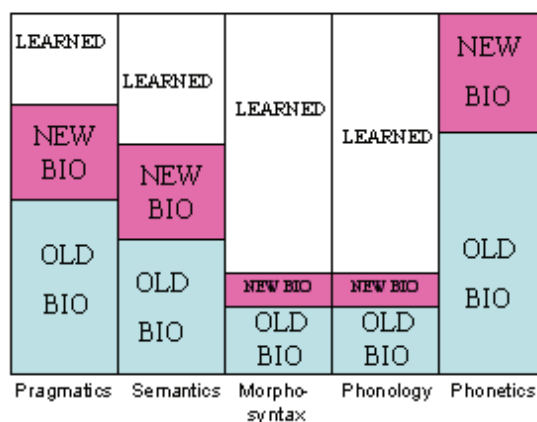


Figure 2. Extremely schematic sketch, not to scale, of the varying antiquity and provenance of subcomponents of a language system.

we cannot contemplate a direct relation between elements of language (words, sentences) and the world. Rather, this relation has to be mediated by minds. This idea is encapsulated in Ogden and Richards' [1923] "triangle of signification" in which it is emphasized that the relation between language and the world is indirect, being mediated by mental entities such as concepts and thoughts. Words and phrases express concepts; the relation of denotation can be reserved for the relation between concepts (mental entities) and things, properties, events and so on in the world.

Now, I assume that some non-linguistic creatures, such as apes and human babies, can have thoughts and concepts. That is, I reject the view that concepts can only be possessed by language-possessing creatures. Some thoughts and concepts pre-existed language. These were private in the creatures that possessed them. With the advent of language, public conventional labels got attached to these private concepts. The attachment of public labels to formerly private concepts had some effect on them. Concepts became socially standardized. But this is to rush ahead. For the moment, suffice it to say that the definition of semantics adopted here is the relationship between concepts (alias thoughts) and entities in the world, as it exists in both linguistic and non-linguistic creatures.

What kinds of concepts and thoughts are non-human animals capable of? It is safe to say that relatively "higher" animals, such as mammals and birds, form mental categories of the types of objects and situations which are relevant to their daily lives, such as different types of food, different food sources and types of predators. They discriminate systematically between such things. This is the evolutionary link to the denotations in an outside world of expressions in human languages. First the concepts of things, events and situations in the world existed

in pre-humans. These were private concepts, about which the creatures did not communicate among themselves. Later, humans became able and willing to attach public labels to these pre-existing concepts, and to use the labels for communication. As a terminological matter, it may be useful to speak of ‘proto-concepts’ in pre-linguistic minds, and to reserve the unmodified term ‘concept’ for those mental entities linked to expressions in a public mode of communication.

Probably some, but not all, of the mental categories non-humans form are learned from experience. There are more or less strong innate dispositions to acquire certain specific categories, and these dispositions are linked to the evolved sensory apparatus of the animal. In some cases, like that of the famous vervet monkey alarm calls ([Seyfarth and Cheney, 1982] — and the similarly specific alarm calls of many other mammalian and avian species) the private concept is linked to a public signal. In these cases, the linkage itself between the concept (e.g. LEOPARD) and the appropriate call (e.g. a ‘bark’) is innately specified to a high degree. The particular behaviour just grows in each animal in a uniform way determined by its genetically driven developmental program.

It is important to make the reservation that a non-human animal has little or no awareness of, or possibility of voluntary control of, its mental categories. A dog cannot, we assume, monitor its mental states to anything like the degree to which adult humans can monitor theirs. Recently claims have been made that non-human animals show some evidence of metacognition, that is, an ability to know their own internal states. For example, an animal trained to make certain categorical distinctions in the lab, can seem to be aware of its own uncertainty in borderline cases [Smith *et al.*, 1995; 1997; 2003a; 2003b; Smith and Washburn, 2005].

It is often said that non-human animals live exclusively in the ‘here and now’. This difference from humans is a matter of degree, and not a categorical difference. For one thing, the concepts of ‘here’ and ‘now’ are extremely flexible. The place referred to as ‘here’ on one occasion may be a dot on a printed page, or it may be this whole universe, as opposed to some alternative or parallel universe, or some region of intermediate size between these extremes. Similarly, ‘now’ can mean this instant, or today, or this century. There is clear evidence that some non-human animals can keep in mind representations of things or events that are not immediately present to their senses. Experiments in ‘object-permanence’ show that a dog, for example, can be aware of the presence of an object hidden behind a screen for up to about five minutes after it last saw it (and without being able to smell it) [Gagnon and Doré 1992; Watson *et al.*, 2001]. A gorilla could recall a sequence of events that it has been shown for up to fifteen minutes after seeing them [Schwartz *et al.*, 2005; 2004]. A chimpanzee who saw food being hidden one day remembered where it had been hidden the next day [Menzel, 2005]. These animals, then, can remember things and events that are not just in the ‘here and now’. Certainly, human abilities far outstrip these non-human performances, but it is a matter of degree. Closely related to this research is work on episodic memory. It has been claimed that only humans have memories for specific events



that have happened to them. Research on several species begins to undermine this suggested absolute difference between humans and non-humans. Scrub jays, for instance, have been shown to remember where they hid food, what kind of food they hid (i.e. perishable or non-perishable) and how long ago they hid it [Clayton *et al.*, 2001; 2003; Clayton and Dickinson, 1998]. Again, such performances are not in the same impressive league as human memories for past events, but the difference is a matter of degree.

It is clear that many animals make plans. Some non-human animals show an ability to distinguish memories of past events from their plans for future behaviour, thus demonstrating an incipient mental distinction between representation of the past from representation of the future. In an experimental situation, rats showed balanced memory limitations for places in a maze that they had already visited and places that they had not yet visited, but presumably planned to [Cook *et al.*, 1983]. This shows the overlap of mechanisms for retrospective memory of past events and prospective memory for planned actions.

Animals attend to individual objects and make categorical judgements about them. The mechanisms of attention, for example involving gaze orientation for visual attention, are distinct from the mechanisms for recognition of objects as belonging to particular categories. An animal's attention is drawn to an object by some simple noticeable change, such as movement or flickering, and subsequently different neural streams feed information from the perceived object into brain areas for categorical recognition, e.g. as food, or as prey. This separation of neural mechanisms, it has been suggested [Hurford, 2003a; 2003b], parallels the logical distinction between a predicate and its argument, where the argument is taken to be an individual variable, not an individual constant. For example, it is suggested that a reasonable schematic representation for what happens in an animal's brain when it spots a wriggling object and then recognizes it as a snake is the formula  $\text{SNAKE}(x)$ . The variable  $x$  stands for the bare object, with no categorical information bound to it; the term  $\text{SNAKE}$  here stands for the categorical predicate-like judgement that the animal makes about the object. More generally, many non-humans animals represent the world in basic ways compatible with the logical schema  $\text{PREDICATE}(x)$ . More thought needs to be given to how complex scenes are represented, and on the overall plausibility of this idea as a pre-human platform for the subsequent evolution of semantic representations in a form something like the formulae of predicate logic. For suggestions, see [Hurford, 2007].

Some remarkable laboratory animals have shown an ability to master second-order judgements, that is to apply predicates to predicates. The late African grey parrot, Alex, could correctly report on the colour, or shape, or material, of an object he was shown [Pepperberg, 2000]. On being shown a red ball, for instance, he could report that it was red, or a ball, whichever superordinate category (colour or shape) he had been asked about. This is the crux of the demonstration that he managed second-order judgements. He was asked, for instance, "What colour is this?", and he would answer "Red". Note that this requires that he knows that

red is a colour. To judge that an object is red is to make a first-order judgement, about a property of the object. To know that red is a colour is to know a second order fact, predicating COLOUR of the predicate RED. Alex, now sadly dead, was trained in a lab. He could generalize across tasks, so there is no question of his simply memorizing rote answers to specific questions. Just what call there might be in a wild state of nature for this advanced ability is not obvious. This illustrates a general point that captive animals often exhibit latent abilities for which they have no apparent need in a wild situation. We will see further examples below. The implication is that the mental abilities of animals, no doubt including humans, are sensitive to their environmental conditions, including social conditions.

Non-human animals do have what can reasonably called thoughts, primarily about the immediate world of perception and action. Human thought, with the aid of language, can be more subtle and complex. When public symbolic labels are attached to previously private (proto-)concepts, their boundaries tend to become sharpened in individual minds and standardized across the community. Much research with children and adults demonstrates that learning distinct labels for different regions in a conceptual space makes discrimination within that space easier [Balaban and Waxman, 1992; Xu, 2002; Booth and Waxman, 2002; Katz, 1963; Goldstone, 1994; 1998].

After the emergence of symbols for relatively basic concepts ('words'), humans at some point began to string them together to encode more complex messages. More complex thoughts could now be held in memory for longer with the aid of 'out loud' rehearsal of the public sequences of words. We are all familiar with how 'thinking out loud' helps us to manage our thoughts. Chomsky tends to the view that this is the primary function of language. It certainly is one function, but the interpersonal communicative function preceded the internal thought-managing function, because the forms that we think in when we think out loud are just those of the language we have learned for communication in our social group. English speakers use English as an instrument for complex thought, Mandarin speakers use Mandarin for the same purpose. The combinatoriality of syntax makes some thoughts accessible which were previously unthinkable. Think, to the extent that you can, of the square root of minus one, or even just minus one. These were concepts inconceivable before combinatorial language.

### *2.1 Pragmatics from an evolutionary point of view*

We, unlike apes, feel free to give each other potentially useful information, and we believe the information given to us. Apes, even domesticated ones such as Kanzi [Savage-Rumbaugh, 1986; 1990; 1999], are notably ungenerous in their communication, though they have learned to be trusting of their human foster parents. By contrast, in the wild, life is much more competitive, and it is unknown for non-human animals to inform each other about the state of the outside world by learned symbols. Some species have evolved small innate systems for such purposes as alerting conspecifics to the presence of predators or food. Vervet monkeys

and honeybees are the best known examples, but there are many others. All such systems in non-human nature are extremely limited in the scope of what they can 'refer' to (e.g. just three types of predator, or the direction and distance of food), and do not need to be learned from experience by the animals concerned.

Non-human animals do, however, communicate. All higher species communicate in some way or other. Here communication is defined as behaving in a way that affects the consequent behaviour of others, other than by straightforwardly causal physical manipulation of their bodies. The most basic, and very widespread, type of communication is purely dyadic, just designed to bring about a response from the receiver of the signal. Courtship behaviour is a central example. The wooing animal has evolved a characteristic way of behaving (e.g. strutting, singing, chest-puffing, distended sexual organs), and wooed animals have evolved complementary ways of responding to such signals. Threat displays such as teeth-baring or pilo-erection, and submissive displays such as cowering and rolling over are further examples. Such behaviour enhances the survival or reproduction chances of the participants and is largely instinctive.

We can see the evolutionary link to human linguistic behaviour in Austin's [1962] phrase 'doing things'. Animals do things to each other in their communication. Humans also use words to do things to each other. A human greeting such as 'Hello' is functionally parallel to a dog's tail-wagging; it is a preliminary move toward subsequent friendly interaction. Of course, human greeting is under voluntary control, whereas the greeting behaviour of dogs is involuntary. Another difference is that human greetings in language are learned arbitrary signals. For each language, you have to learn a different conventional greeting word or phrase. But the functional connection to animal behaviour remains. Most communication in language is not purely dyadic like a 'Hello' greeting. Almost all linguistic communication is referential, in the sense of being about something other than the speaker or hearer. But the 'doing-things-to-each-other' aspect of communication is always present. Why else would we speak? In linguistic pragmatics, this aspect of language is captured by the term 'illocution'. The illocution of an utterance is what is done, typically to the other person, in making that utterance. For instance, my uttering "the door is open" can be used to invite you, to dismiss you, to warn you of danger, or to get you to close the door, depending on the mutually understood context.

Mutual understanding of the purposes of communication is omnipresent in human linguistic behaviour. When someone says something, we assume it is said for a reason, and we try to divine the speaker's goal [Sperber and Wilson, 1986]. Sometimes the process of figuring out a speaker's actual intent can be quite circuitous, as in my example of 'the door is open'. We humans do this with the benefit of a well developed 'Theory of Mind'. We know the range of possible psychological states that a speaker may be in, and we can guess quite well what the speaker in a given situation knows and does not know about that situation. Much human discourse is consequently oblique. A typical exchange might be "There's no milk", followed by "It's Sunday". Such oblique communication works because the interlocutors

understand each other's possible motives and current knowledge. Non-human animals also display some very basic understanding of the moods and knowledge of others. A chimpanzee, for example, can tell the difference between very similar physical actions, according to whether they are aggressive (e.g. teasing) or unsuccessfully cooperative (e.g. fumbling) [Call *et al.*, 2004]. Other experiments seem to show that a chimpanzee can know whether another, dominant, chimpanzee has seen, and thus knows about, a particular food item [Call and Tomasello, 2005; Hare *et al.*, 2000; 2001]. Humans far outstrip non-humans in this 'mind-reading' behaviour, but the difference is a matter of degree.

The evolutionary move from merely dyadic communication, involving only the sender and the recipient, to triadic signalling, where messages are also about some other object or event in the world, is facilitated by a capacity for joint attention. In simple cases, when humans converse about some object in the context of their encounter, both parties attend to the object. Uttering "Pass me that cup" assumes in some sense that both speaker and hearer can attend to the cup in question. Some non-human animals are adept at following the gaze of others, thus bringing about joint attention to the same object [Bräuer *et al.*, 2005; Call *et al.*, 1998].

If for some reason language is not available, as when people don't speak the same language, humans use pointing to achieve joint attention. Although they are physically capable of pointing, apes in the wild never use any form of pointing to draw attention to objects [Leavens, 2004; Leavens and Hopkins, 1998]. By contrast, human toddlers often voluntarily point to objects, apparently merely to draw attention to them for the purpose of sharing an interest with an adult. In captivity, apes have learned to point to things, but almost exclusively as a means of requesting objects. Compared to human babies, apes are mercenary or instrumental in their dealings with humans and with other apes. Any pointing is motivated by directly selfish ends. Children are not like that.

This raises a central puzzle in the evolution of language. Why should any creature voluntarily share information with another? Information can be valuable, and a selfish disposition advocates that one should keep valuable information to oneself. Various theories developed in biology begin to unravel this puzzle. Passing information to one's close kin (e.g. offspring or siblings) can enhance the fitness of individuals with whom one shares genes, and thus the sharing of information is expected to be adaptive between close kin, by a process known as kin selection [Hamilton, 1964]. Many non-human animals act altruistically toward close kin, and even humans have been shown to share information more with close kin than with unrelated individuals. This establishes a parallel between humans and non-humans. But humans willingly share information, and more generally act altruistically, with non-kin. Here the theory of reciprocal altruism [Trivers, 1971] can make some headway. Within a social group, as theoretical models (computational and mathematical) models show, "tit-for-tat" behaviour is adaptive. This is, a strategy of acting altruistically toward another individual is advantageous if there is some reasonable assurance that the altruism will be reciprocated at some future time. There is evidence for such reciprocal altruism in some non-human species,

as when chimpanzees form alliances for mutual defence, and in some food-sharing activity [de Waal, 1989]. Reciprocal altruism is much better developed in human communities. We are more disposed to communicate cooperatively with people in our own social group than with outsiders, and within-group cooperation is typically reciprocated. A further motivating factor in human signalling of valuable information is that it brings prestige to the communicator, and is thus adaptive [Dessalles, 1998].

In short, human communication in language shares with animal communication the doing-things-to-each-other feature; many non-human species have limited instinctive unlearned systems for alerting others to things crucial to their survival such as predators or food; non-human animals show hints of what it takes to figure out the communicative intentions of others, such as gaze-following and a rudimentary Theory of Mind, but in the wild they do not apply these abilities in learned systems for communicating referentially about a wide range of external objects and events. The difference lies largely in the degree of natural cooperation that is built into the genes and the societies of humans and non-humans. We humans (believe it or not) are the species most disposed to act altruistically and cooperatively with members of our own social group.

### 3 PHONETICS FROM AN EVOLUTIONARY POINT OF VIEW

Phonetics is conveniently defined from an evolutionary viewpoint as the hardware of speech production and perception. Although human language exists in both oral and manual modalities, it does not seem (on the evidence so far) that human manual dexterity is specially adapted for signing, or that our vision is specially adapted for interpreting manual signs. On the other hand, the output machinery for speech, namely the whole human vocal tract, is clearly adapted, probably rather recently, for speech. As for the input stream, there is less agreement about whether human hearing is specially adapted for speech processing. I will discuss the question of human hearing first.

Mammalian hearing, to the cochlea, is rather uniform. In experimental situations, it can be shown that chinchillas have similar categorical perception of voice onset time (e.g. the difference between a [b] and a [p]) as humans. Tamarin monkeys make the same discriminatory judgements of rhythm in different languages (e.g. the rhythmic difference between Dutch and Japanese) as human babies [Tincoff *et al.*, 2005]. Chimpanzees perceive the differences between simple syllables (e.g. [ba], [ga], [da]) in the same way as humans [Kojima *et al.*, 1989]. And chimpanzees can do vocal tract normalization, that is they can recognize the ‘same sound’ spoken by different speakers [Kojima and Kiritani, 1989]. Opposing such evidence proposed against any special adaptation for speech in human hearing, the point has been made that normal speech perception by humans involves putting all such separate abilities together very fast in extracting complex hierarchical meanings from the stream of speech, and there is no evidence that non-human animals can manage that [Pinker and Jackendoff, 2005]. The bonobo Kanzi can

follow simple spoken instructions such as “put the coffee in the milk”, so evidently he can pick individual words out of the stream of speech [Savage-Rumbaugh *et al.*, 1993]. On the other hand, it has also been shown that chimpanzees’ auditory working memory is impoverished as compared to humans [Hashiya and Kojima, 2001].

The issue of whether human hearing is specially adapted for speech is distinct from the issue of whether humans have distinct mechanisms for processing speech sounds and other environmental sounds (such as the sound of wind blowing or rocks falling). Humans do have mechanisms for speech processing that are separate from their mechanisms for processing other sounds [Lieberman and Mattingley, 1989]. At the periphery of the system there is no difference, but at some point in the processing system there is a filter that directs speech sounds to brain regions specialized for speech processing, not surprisingly largely in the left hemisphere. But this dual-system arrangement is not special to humans. Many animals, including primates, have at least partly separated brain mechanisms for processing the calls of conspecifics and other environmental noises [Zoloth *et al.*, 1979; Heffner and Heffner, 1984; 1986; Ghazanfar and Hauser, 2001; Hauser and Andersson, 1994; Ghazanfar *et al.*, 2001]. Within humans, the slogan “speech is special” applies, because of this separation between speech sounds and other sounds. But it does not follow that humans are special in this regard, because as we have seen, many primates also have distinct mechanisms from processing the communicative sounds of their species.

Coming now to human speech production, there is no doubt that specialized adaptations have occurred in our species fairly recently. All of our speech apparatus has been exapted from other functions. The tongue and teeth originally evolved for eating, the lungs for breathing, and the glottis (vocal cords) for keeping water out of the lungs and bracing the chest at times of exertion.

The most widely discussed adaptation for speech is the lowering of the larynx. In all other mammals the normal position of the larynx is close up behind where the nasal passage joins the oral passage, just behind the velum. This is also the position of the larynx in newborn human infants, which allows them to breathe and suckle at the same time. During the first half year of life the human larynx lowers to near its later adult position. In this way ontogeny recapitulates phylogeny, as the adult human larynx has lowered in our evolution from apes. A lowered larynx creates a two-chamber supraglottal vocal tract. The rear chamber, the pharynx, and the front chamber, the mouth, can be narrowed or broadened complementarily. As a result, the vibrating air column used in vowel production can either pass first through a narrow tube and later through a wider tube (giving an [a] vowel), or first through a wider tube and then through a narrower tube (giving an [i] vowel). This flexibility of the upper vocal tract make possible a range of different vowel qualities that apes cannot produce. It seems reasonable that this is an adaptation allowing for a greater variety of spoken signals [Lieberman, 1984].

The story of larynx lowering is slightly more complicated. Some non-humans can lower their larynx dynamically, as dogs do momentarily when barking, and male

deer do in a more stately fashion when roaring [Fitch and Reby, 2001] But such animals nevertheless do not have the two-chamber vocal tract that makes possible the range of human vowel qualities. Although the selective value of a lowered larynx was largely for greater versatility in vowel production, the further slight lowering of the larynx in human males at puberty is also probably an adaptation for sexual selection. The difference between adult male and female voices is by far the most prominent case of sexual dimorphism in language.

Another recent adaptation is voluntary control of vocalization and breathing. Other apes have little or no voluntary control over their vocalizations. There are no significant connections from cortex to the larynx. Ape cries are spontaneous and automatic. Humans also have good control over their breathing. During speech, an outbreath may last up to thirty seconds, with the air being released in a very slow and controlled manner. Comparisons of the skeletal channels for the nerves that work the muscles involved in breathing shows a recent expansion in humans, suggesting an adaptation for greater control of breathing [MacLarnon and Hewitt, 1999]. A similar claim for the hole in the base of the skull through which the nerves controlling the tongue pass has not been substantiated [Kay *et al.*, 1999; DeGusta *et al.*, 1999], though there is little doubt that humans have finer control over the configurations of their tongues than other apes. Human speech production is exquisitely orchestrated, and the human vocal tract and the cerebral machinery controlling it are undoubtedly recent adaptations since divergence from our last common ancestor with the chimpanzees about six million years ago.

#### 4 PHONOLOGY FROM AN EVOLUTIONARY POINT OF VIEW

Phonology is defined as the patterns of speech in languages. Languages organize their sound patterns within the possibilities afforded by the auditory and vocal apparatus. The physical apparatus is pretty much universal, give or take some individual variation not reflected in the organization of particular languages. The raw material of phonological organization is given by the mobility of the jaws and velum, the great flexibility of the lips and tongue and the several possible states of the glottis while air passes through it. The range of possible noises that can be made using these instruments is vast. Imagine a sequence of random uncoordinated impulses to these articulators. The product would be nothing like speech. Speech is to such random vocal noise as ballet is to the uncoordinated staggering and falling of a drunkard. Speech in languages is disciplined into repertoires of precisely specified and tightly controlled conventional moves. Acquiring perfect pronunciation in a language requires early exposure and practice. People starting a new language after the age of about eight rarely achieve perfect pronunciation.

The vocal articulators are like an orchestra [Browman and Goldstein, 1992]. During tuning, each instrument acts independently of the others, and the result is cacophony. For example, the lips can open and close at any time, the vibration of the vocal cords can be switched on or off at any time, the tongue can move between any of its possible configurations at any pace and the velum can be raised

or lowered at any time. Nothing in the inborn physical apparatus dictates that any of these actions be coordinated. All spoken languages, however, are structured in terms of basic building blocks, namely phonetic segments and syllables, which are produced by strict coordination of the actions of the various articulators. Without such coordination, speech sounds as they are commonly understood, and for which the International Phonetic Alphabet has symbols, do not exist.

The basic training of the speech apparatus to produce these discrete speech sounds occurs during a child's development. Human infants, unlike the young of other apes, spontaneously babble, exercising their vocal apparatus at first in random ways but progressing toward sequences which more recognizably consist of speech sounds organized into syllables of consonant-vowel (CV) structure. A basic alternation between consonants and vowels makes each individual sound easier to recognize as a self-standing unit. The CV structure is found in all languages. Some languages have developed more complex syllable structures, with short clusters of consonants and more complex vowels (e.g. diphthongs), but any tendency toward such complexity is at the cost of easy speech perception. The auditory feedback received by the babbling infant helps it to map its motor movements onto acoustic patterns. The disposition to babble is thus adaptive in a social group that already benefits from communication in speech. It seems likely that a capacity for finer tuning of the articulators and more precise coordination of their interaction evolved biologically as the benefits of well articulated speech emerged. This would have been a case of gene-culture (more specifically gene-language) co-evolution.

We analyze languages as having inventories of phonemes just because these units are re-used over and over in many different words. Given a massive vocabulary of tens of thousands of words, it is costly for each separate word form to be phonetically *sui generis*, memorized holistically. In every language there is a handful of expressive forms that resist representation as a sequence of the normal phonemes of the language. Examples include: the expression of disgust conventionally, but inaccurately, spelled as 'Ugh'; the alveolar click used to express disapproval (with 'tsk' as an attempted spelling); the bilabial affricate used to respond to cold (spelled 'brrr'), and so on. Such expressions are not composed of phonemes in regular use elsewhere in the language. This type of expression is perhaps an evolutionary remnant of a pre-phonological stage when speech was limited and not organized around a versatile inventory of re-usable phonemes. But once large vocabularies became available it was not practical to organize the bulk of the word store in this way. The process by which re-usable phonological units get crystallized out of a mass of inchoate vocalizations has been modelled computationally [Zuidema and de Boer, 2009; de Boer and Zuidema, 2010]. The competing adaptive pressures leading to the emergence of small inventories of systematically re-usable segments are ease of articulation and mutual distinctiveness of words from each other. This evolutionary process can be seen as an instance of self-organization of a system in the environment provided by the phonetic apparatus and given the twin pressures just mentioned.

Self-organization can also be seen in the evolution of vowel inventories. Mod-



elling vowels is relatively straightforward as the continuous articulatory and acoustic spaces that they occupy are well understood, with only three main dimensions that do most of the work. Languages differ in the number of their vowel phonemes, from as few as two to over a dozen as in English. In the statistical distribution of the size of vowel inventories, the most common size is five vowels, roughly [i], [e], [a], [o], [u], as in Spanish. Systems with fewer than five vowels and with more than five vowels are decreasingly common in languages as the number differs from five. However many vowels a language has, they tend to be arranged symmetrically around the vowel space, this making maximum use of the space. The evolution by self-organization of vowels from randomly distributed beginnings has been simulated computationally. The model captures well the distribution of different numbers of vowels across languages. The model can be interpreted as mimicking the ancient processes by which well-organized vowels systems emerged in the earliest languages. The joint adaptive pressures causing this emergence are ease of articulation and mutual distinctiveness of each vowel from all the others. It is these same pressures that maintain vowel systems in extant languages in roughly the same symmetrical states over the course of their histories.

The studies surveyed above account quite successfully for the gross features of the phonological organization of all languages, namely their basic CV structure, their basis in sets of consonant and vowel phonemes, and the typical distribution of vowels in the acoustic/articulatory space. Modelling has not yet progressed to the fine detail of the ways in which adjacent sounds in a language affect each other, though this is a pervasive aspect of phonological organization. But we can nevertheless see an evolutionary dimension in such phonological effects. Natural phonetic influences which are felt by all speakers, modifying the canonical form of a phoneme, can become conventionalized, so that a synchronic phonological rule describes the regular effect. For instance, it is natural for a canonically voiced phoneme to be devoiced in anticipation of a following pause (as pauses are voiceless). In German, this devoicing has become institutionalized and extended to all word-final canonically voiced phonemes. We can see the modern synchronic rule as the trace of more optional processes earlier in the history of the language. Many synchronic phonological rules are the lasting after-effects of earlier historical sound changes in a language.

## 5 SYNTAX FROM AN EVOLUTIONARY POINT OF VIEW

As mentioned in the introduction, the two major contenders for crucial evolutionary changes leading to modern language-using humans have been (1) a capacity for complex syntax, and (2) a capacity to learn tens of thousands of arbitrary symbols. The former, the capacity for syntax, has always been regarded as the most challenging and theoretically interesting. A large memory for symbols was regarded as less interesting. The exciting focus of linguistic theory was syntax. Humans obviously have a unique capacity for syntax. From the early days of generative grammar in the 1950s until the mid-90s, it was assumed that this capacity

was complex, comprising up to half a dozen interacting principles. These principles were assumed to be innate, not needing to be learned, and arbitrary, not motivated by functional factors. A child learning a complex human language was assumed to receive substantial help from inborn knowledge of the abstract ways in which languages work. Here ‘abstract’ means that the innate principles were held to deal in terms of generalizations over syntactic categories (such as Noun, Noun Phrase, Verb, Verb Phrase), and general constraints on operations on the hierarchical tree structures of sentences (for example, an element could not ‘move’ over certain specified constituents). Discovering the set of these inborn principles, and the manner of their interaction, was the central goal of generative syntactic theory. Theorists in the generative paradigm became known as ‘formalists’. Outside this paradigm, the ‘functionalists’ objected to the generativists’ emphasis on abstraction and their lack of concern for functional explanations of the properties of language.

In the mid-1990s a major revision of generative syntactic theory appeared in the form of Chomsky’s ‘Minimalist Program’ [Chomsky, 1995]. Here the number of innate principles was in theory reduced to just one. It was suggested that perhaps the only distinctive feature of the human syntactic capacity is a capacity for recursively combining words and the phrases they compose [Hauser *et al.*, 2002]. The central operation of syntax was ‘Merge’. Since even simple operations, if applied recursively, can lead to impressive complex structures (and, sociologically, because of old habits), the discussions of adherents to the Minimalist Program continued to have a highly abstract flavour. It became clear, however, that there was evolving convergence, from many camps, on the simple idea that what is distinctive about the human syntactic capacity is just semantically compositional combinatoriality. Various generative, but non-Chomskyan, theoretical frameworks, such as Head-driven Phrase Structure Grammar (HPSG) [Pollard and Sag, 1987; 1994; Levine and Meurers, 2006] and Construction Grammar [Fillmore and Kay, 1993; Fillmore *et al.*, 2003; Goldberg, 1995; 2006; Croft, 2001], had already been pointing in this direction for several decades. From an evolutionary point of view, the reduced complexity of the syntactic apparatus innately programmed to develop in the child was welcome, as it simplified the likely course of human evolution. The evolution of one trait is less challenging to explain than the evolution of several mutually influencing traits. Biologists interested in human evolution welcomed this theoretical development in linguistics. Nevertheless, even with this simplification, it was still thought that there had been in human evolution a qualitative leap from non-syntactic ‘protolanguage’ to fully combinatorial language. No continuity was seen between unstructured stringing together of words and the more complex morphosyntactic systems seen in modern languages. In box diagrams of the architecture of language, separate boxes for ‘lexicon’ and ‘syntax’ were assumed. At a certain level of granularity this is acceptable. Analogously, any sensible description of human anatomy identifies separate organs. The challenge to evolutionary theory is to explain, for example, how the mammalian eye could have evolved from a non-eye, or a backbone from a non-backbone. How, without a biologically

implausible saltation, could human syntax have evolved out of non-syntax?

Recently, a way of approaching this question has emerged, mainly under the banner of Construction Grammar, and with support from much research in child language development (e.g. [Bates and Goodman, 1997]). It is suggested that there is a ‘lexicon-syntax continuum’. The lexicon can contain items of varying complexity, from simple words to whole memorized sentences (or perhaps even the whole memorized Koran). Many conventional idioms and proverbs are stored as wholes, rather than being productively generated. All such stored items are known as ‘constructions’; a word is a construction, a whole memorized sentence is a construction. Crucially, constructions may also vary in flexibility or abstractness. A certain idiom, for example, may not be completely rigidly specified, but may appear in different permutations. The idiom *kick the bucket* can be modified for past or non-past tense, so we can have both *kicks the bucket* and *kicked the bucket*. The idiom is stored as a whole, but with a variable slot for specification of tense. Somewhat more flexible are ‘syntactic idioms’ such as *take advantage of*. In this construction, the verb *take* and the noun *advantage* are in a constant verb-object syntactic relationship, and this can interact with other constructions, as in the passive *Advantage was taken of John* or *John was taken advantage of*. Putting it briefly, humans have evolved a capacity for storing building blocks of various sizes and of varying degrees of flexibility, and a capacity for combining them with others. The first building blocks ever used were small and inflexible. Later, somewhat larger and/or more flexible units were invented and re-used if they proved useful. The initial step from non-combining to combining is still an inevitable qualitative shift, but it did not immediately give rise to an explosion into the extreme productivity of modern languages. In a story of how complex human languages emerged, it is plausible that the very earliest combinations were of the simplest and least abstract items, like the *Me Tarzan* and *You Jane* of the movies. The evolution from that early stage to modern complex languages was a gradual co-evolutionary process, involving cultural invention and re-use of progressively more complex and flexible stored forms, accompanied by biological expansion of the available mental storage space and speeding-up of the possible online combinatorial processes.

Thinking of syntax from an evolutionary point of view prompts a revision of a central tenet of generative theory, namely the relationship between competence and performance.

Competence is the specifically grammatical knowledge in a speaker’s head that allows him to produce and interpret complex sentences. Competence in a language is always there in the speaker’s head, whether it is being used or not. It is thus ‘timeless’. Performance, on the other hand, is the actual processes, located in time and space, of production and interpretation. Some such distinction is indispensable, like the distinction between a computer program and its running at different times, with different inputs and outputs. The focus of generative theory has always been on competence, with performance factors such as limitations on memory and speed of processing being relegated to the status of distracting noise. A central

example is the case of centre-embedded clauses, which I will explain with some examples. Separately, all the following expressions are grammatical sentences: *The mouse the cat caught died*; *The cat the dog chased escaped*; *The dog the man kicked howled*; *The man I saw laughed*. They can be acceptably combined to a certain extent, so that *The mouse the cat the dog chased caught died* can, with suitable context and intonation, just about be understood. Nevertheless this last sentence clearly puts strain on the language processor. Further combinations, such as *The mouse the cat the dog the man I saw kicked chased caught died* are impossible to process without paper and pencil, or concentration on the written form. Generative theory has always held that since such complex examples are formed by exactly the same rules as simpler examples, they must be within a speaker's competence, though it is admitted that they are outside the limits of his performance. In short, generative theory has resisted any quantitative or numerical element in competence. Quantitative limitations belong to a theory of performance, not to a theory of competence. This is a coherent and understandable theoretical position. But from an evolutionary point of view, it is not possible to see how a capacity to acquire competence in a language can ever have been separate from a capacity for production and interpretation of the objects defined by that competence. Twin interdependent capacities, for internal representation of the permissible regularities of a language (competence), and for putting that knowledge to use on specific occasions (performance) must have co-evolved. Competence without a capacity for performance would have had no impact on the world, so no evolutionary advantage, and complex performance could not happen without a complex internal program to guide it. Hurford [2011] develops this idea with a construct of 'competence-plus', a package of rule-like representations combined with numerical limits on their application, for instance limits in depth of embedding. In the evolution of mankind, there was parallel linked growth of the possible complexity of internal representations of the regularities of a language and the quantitative limits on what could be produced or interpreted.

Larger and more flexible constructions can be advantageous to their users, both speakers and hearers, allowing more versatile and less clumsy communication. Complex syntax is especially advantageous when it is compositionally linked to semantics, that is when the meaning of a complex expression is a function of the meanings of the parts and the way these parts are put together. Complex syntax is easier to process when one is able to interpret it as meaningful. Human parsing of complex sentences is a process of deciphering the sequence of words into a representation of some propositional content, plus some indication of the pragmatic intent of the speaker. Parsing uses clues from the words themselves, from markers of grammatical structure in the sentence, and from the whole situational context of the utterance. Long paradoxical or nonsensical strings of words are less easy to parse than meaningful ones of the same length. The evolutionary driver of the modern human capacity for complex syntax was surely the semantic carrying power of complex sentences.

In the history of modern languages we see a process of 'grammaticalization'

[Heine and Kuteva, 2007; Hopper and Traugott, 1993; Givón, 2009]. Unless contaminated by contact with other languages, there is a tendency, in all languages, for common patterns in discourse to become entrenched, or conventionally fixed, not just as possible ways of expressing certain meanings, but as required ways of expressing meanings. In English, for instance, every non-imperative sentence must have a subject. A string such as just *came today*, with no overt subject, is not a full or proper sentence, although it could be perfectly understandable in an appropriate discourse. In other languages, this is not a requirement, as for example *e venuto oggi* is grammatical in Italian. But even in languages like Italian, so-called null-subject languages, there is a grammatical indication of an understood subject, in the form of the agreement of the verb. It is widely held that grammatical subjects in languages are fossilized Topics. Some form of Topic-Comment structure is universal in languages. All languages have a way of marking the expression denoting the thing that is being talked about (the Topic of a sentence), as opposed to what is being said about it (the Comment, or Focus). Some languages do not mark the subjects of sentences at all, and some only mark them optionally with few specific markers. In English and many other languages, verbal agreement singles out the subject of a sentence. In many languages the relationship between agreement inflections on verbs and subject pronouns is transparent, suggesting a diachronic process of grammaticalization of subjects from pronouns in an overtly Topic-Comment structure, as in *That guy, he's crazy* [Givón, 1976].

I have singled out the grammatical role of subject because it is a centrally grammatical notion, as opposed to a semantic notion like Agent (the 'doer of the action') or a pragmatic notion like Topic (what a speaker assumes is shared information). The grammatical role of subject has emerged, by grammaticalization from the non-grammatical discourse-structural function of Topic repeatedly and independently in the histories of many languages. Many other widespread aspects of the grammars of modern languages, such as their typical inventories of syntactic categories ('parts of speech') have also arisen through grammaticalization. Heine and Kuteva [2007] survey a wide range of languages and give many examples of particular parts of speech and grammatical markers arising historically from other word-classes. For instance, prepositions commonly arise from bodypart nouns, such as *back*; auxiliary verbs arise from main verbs (e.g. *have*); relative clause markers often derive from demonstratives (e.g. English *that*) or from question words (e.g. *which*). The process of grammaticalization is overwhelmingly unidirectional, and so it is plausible to reconstruct earlier stages of human languages as lacking the grammatical features produced historically by grammaticalization. Indeed, the few modern cases which come close to genuine language creation *de novo*, such as Nicaraguan Sign Language, show such bare featureless properties in their early stages, with no or few grammatical markers. Very quickly, however, grammaticalization processes kick in and languages soon develop grammatical structure characteristic of modern languages. The grammatical complexity of modern languages is a historical product, enabled, to be sure, by a biologically endowed capacity to manage such complex systems with facility.

Some degree of syntactic complexity exists in nature without any compositional syntax. The songs of some whales [Payne and McVay, 1971] and many songbirds (see, e.g., Todt and Hultsch, 1996; 1998]) are hierarchically structured into what can naturally be called phrases, but these ‘phrases’ make no meaningful contribution to the overall meaning of the complex song. As far as we know, an entire complex birdsong functions either as an invitation to mate or as a warning to keep away from the singer’s territory. Birdsong is syntactically complex, but carries no meaning that is a function of the meanings of its constituent notes and phrases. Indeed, the notes and phrases have no meanings. It has been suggested, by thinkers as illustrious as Rousseau, Darwin and Otto Jespersen, that pre-humans possessed some capacity for such syntactically complex song before it became a vehicle for the expression of messages composed from the meanings of the parts. This is possible, but a problem with the story is that we find no such complex syntactic behaviour in species closely related to humans, in particular in apes and almost all primates, with the possible exception of gibbons.

Some monkeys and apes do string a few meaningful elements together to make sequences that are also meaningful, but the meanings of the whole strings are apparently not a function of the meanings of the parts. For instance, a species of monkey observed in the wild by Arnold and Zuberbühler [2006] has two alarm calls, one for eagles and one for leopards. A combination of these two calls seems to function as a summons for, or comment on, unpanicky movement of the group to another location. It is not clear that the meaning of the two-element combination (roughly ‘all move’) is a function of the meanings of the two parts (roughly ‘eagle’ and ‘leopard’). Truly semantically compositional syntax occurs only in humans, and humans have taken it to a high order of complexity.

Summarizing the evolutionary view of language structure, the human language capacity, especially the capacity for massive storage of constructions large and small, with greater or lesser flexibility and combinability, and the facility for recursively combining constructions fast during speech production, and disentangling them fast during speech perception, were selected because of the advantages of carrying propositional information. Speakers capable of greater fluency benefitted individually, by gaining prestige. Groups containing such speakers, and hearers capable of understanding them, prospered because of the advantages of communicating informatively. The complex structures of individual languages evolved historically over many millennia through such processes as the self-organization we have seen in phonology and grammaticalization in syntax. An evolutionary approach to the language faculty and to languages asks ‘How did they get to be that way?’ I hope to have shown that there are some answers well worth considering.

## BIBLIOGRAPHY

- [Arnold and Zuberbühler, 2006] K. Arnold and K. Zuberbühler. Language evolution: Semantic combinations in primate calls. *Nature* 441, 303, 2006.
- [Austin, 1962] J. L. Austin. *How to Do Things with Words*. Cambridge, MA: Harvard University Press, 1962.

- [Axelrod, 1984] R. Axelrod. *The Evolution of Cooperation*. New York: Basic Books, 1984.
- [Axelrod, 1997] R. Axelrod. *The Complexity of Co-operation*. Princeton, NJ: Princeton University Press, 1997.
- [Balaban and Waxman, 1992] M. Balaban and S. Waxman. Words may facilitate categorization in 9-month-old infants. *Journal of Experimental Child Psychology* 64, 3–26, 1992.
- [Bates and Goodman, 1997] E. Bates and J. C. Goodman. On the inseparability of grammar from the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes* 12(5/6), 507–584, 1997.
- [Booth and Waxman, 2002] A. E. Booth and S. R. Waxman. Object names and object functions serve as cues to categories in infancy. *Developmental Psychology* 38(6), 948–957, 2002.
- [Bräuer *et al.*, 2005] J. Bräuer, J. Call, and M. Tomasello. All great ape species follow gaze to distant locations and around barriers. *Journal of Comparative Psychology* 119(2), 145–154, 2005.
- [Browman and Goldstein, 1992] C. P. Browman and L. Goldstein. Articulatory Phonology: An Overview. *Phonetica*, 49, 155–180, 1992.
- [Bybee, 1985] J. L. Bybee. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins, 1985.
- [Bybee *et al.*, 1994] J. L. Bybee, R. Perkins and W. Pagliuca. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press, 1994.
- [Call *et al.*, 1998] J. Call, B. Hare, and M. Tomasello (1998). Chimpanzee gaze following in an object choice task. *Animal Cognition* 1, 89–99, 1998.
- [Call and Tomasello, 2005] J. Call and M. Tomasello. What chimpanzees know about seeing, revisited: an explanation of the third kind. In N. Eilan, C. Hoerl, T. McCormack, and J. Roessler (eds.), *Joint Attention: Communication and Other Minds*, pp. 45–64. Oxford: Oxford University Press, 2005.
- [Call *et al.*, 2004] J. Call, B. Hare, M. Carpenter, and M. Tomasello. ‘Unwilling’ versus ‘unable’: chimpanzees’ understanding of human intentional action. *Developmental Science* 7, 488–498, 2004.
- [Chomsky, 1965] N. A. Chomsky. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.
- [Chomsky, 1975] N. A. Chomsky. *The Logical Structure of Linguistic Theory*. New York: Plenum, 1975.
- [Chomsky, 1981] N. A. Chomsky. *Lectures on Government and Binding: the Pisa Lectures*. Dordrecht, Netherlands: Foris, 1981.
- [Chomsky, 1995] N. A. Chomsky. *The Minimalist Program*. Number 28 in Current Studies in Linguistics. Cambridge, MA: MIT Press, 1995.
- [Clayton *et al.*, 2003] N. S. Clayton, T. J. Bussey, N. J. Emery, and A. Dickinson. Prometheus to Proust: the case for behavioural criteria for mental ‘time travel’. *Trends in Cognitive Sciences* 7(10), 436–437, 2003.
- [Clayton and Dickinson, 1998] N. S. Clayton and A. Dickinson. Episodic-like memory during cache recovery by scrub jays. *Nature* 395, 272–274, 1998.
- [Clayton *et al.*, 2001] N. S. Clayton, D. Griffiths, N. Emery, and A. Dickinson. Elements of episodic-like memory in animals. *Philosophical Transactions of the Royal Society of London B* 356(1413), 1483–1491, 2001.
- [Cook *et al.*, 1983] R. Cook, M. Brown, and D. Riley. Flexible memory processing by rats: use of prospective and retrospective information. *Journal of Experimental Psychology: Animal Behavior Processes* 11, 453–469, 1983.
- [Croft, 2001] W. Croft. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press, 2001.
- [de Boer and Zuidema, 2010] B. de Boer and W. Zuidema. Multi-agent simulations of the evolution of combinatorial phonology. *Adaptive Behavior*, 18(2), 141–154, 2010.
- [DeGusta *et al.*, 1999] D. DeGusta, W. H. Gilbert and S. P. Turner. Hypoglossal canal size and hominid speech. *Proceedings of the National Academy of Sciences of the U.S.A.*, 96(4), 1800–1804, 1999.
- [Dessalles, 1998] J.-L. Dessalles. Altruism, status and the origin of relevance. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (eds.), *Approaches to the Evolution of Language*, pp. 130–147. Cambridge: Cambridge University Press, 1998.

- [de Waal, 1989] F. B. M. de Waal. Food sharing and reciprocal obligations among chimpanzees. *Journal of Human Evolution* 18, 433–459, 1989.
- [Dobzhansky, 1973] T. Dobzhansky. Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher* 35, 125–129, 1973.
- [Fillmore and Kay, 1993] C. Fillmore and P. Kay. *Construction Grammar Coursebook*. Berkeley, CA: Department of Linguistics, University of California, Berkeley, 1993.
- [Fillmore et al., 2003] C. Fillmore, P. Kay, L. A. Michaelis, and I. Sag. *Construction Grammar*. Chicago: University of Chicago Press, 2003.
- [Fitch and Reby, 2001] w. T. Fitch and D. Reby. The descended larynx is not uniquely human. *Proceedings of the Royal Society, London*. 268, 1669–1675, 2001.
- [Gagnon and Doré, 1992] S. Gagnon and F. Doré. Search behavior in various breeds of adult dogs (*Canis familiaris*): object permanence and olfactory cues. *Journal of Comparative Psychology* 106, 58–68, 1992.
- [Ghazanfar and Hauser, 2001] A. A. Ghazanfar and M. D. Hauser. The auditory behaviour of primates: A neuroethological perspective. *Current Opinion in Neurobiology* 11, 712–720, 2001.
- [Ghazanfar et al., 2001] A. A. Ghazanfar, D. Smith-Rohrberg, and M. D. Hauser. The role of temporal cues in rhesus monkey vocal recognition: Orienting asymmetries to reversed calls. *Brain, Behavior and Evolution* 58, 163–172, 2001.
- [Givón, 1976] T. Givón. Topic, pronoun and grammatical agreement. In C. N. Li (Ed.), *Subject and Topic*, pp. 149–188. New York: Academic Press, 1976.
- [Givón, 1979] T. Givón. *On Understanding Grammar*. New York: Academic Press, 1979.
- [Givón, 1990] T. Givón. *Syntax: A Functional-Typological Introduction*. Amsterdam: John Benjamins, 1990.
- [Givón, 2009] T. Givón. *The Genesis of Syntactic Complexity*. Amsterdam: John Benjamins, 2009.
- [Goldberg, 1995] A. E. Goldberg. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press, 1995.
- [Goldberg, 2006] A. E. Goldberg. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press, 2006.
- [Goldstone, 1994] R. L. Goldstone. Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General* 123, 178–200, 1994.
- [Goldstone, 1998] R. L. Goldstone. Perceptual learning. *Annual Review of Psychology* 49, 585–612, 1998.
- [Hamilton, 1964] W. D. Hamilton. The genetical evolution of social behavior. *Journal of Theoretical Biology* 7, 1–16, 17–52. Parts I and II, 1964.
- [Hare et al., 2000] B. Hare, J. Call, B. Agnetta, and M. Tomasello (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour* 59, 771–785, 2000.
- [Hare et al., 2001] B. Hare, J. Call, and M. Tomasello (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour* 61(1), 139–151, 2001.
- [Hashiya and Kojima, 2001] K. Hashiya and S. Kojima. Hearing and auditory-visual intermodal recognition in the chimpanzee. In T. Matsuzawa (Ed.), *Primate Origins of Human Cognition and Behavior*, pp. 155–189. Tokyo: Springer Verlag, 2001.
- [Hauser and Andersson, 1994] M. D. Hauser and K. Andersson. Left hemisphere dominance for processing vocalizations in adult, but not infant, rhesus monkeys: Field experiments. *Proceedings of the National Academy of Sciences of the U.S.A.* 91, 3946–3948, 1994.
- [Hauser et al., 2002] M. D. Hauser, N. Chomsky, and W. T. Fitch. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579, 2002.
- [Heffner and Heffner, 1984] H. E. Heffner and R. S. Heffner. Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science* 4670, 75–76, 1984.
- [Heffner and Heffner, 1986] H. E. Heffner and R. S. Heffner. Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by Japanese macaques. *Journal of Neurophysiology* 56, 683–701, 1986.
- [Heine and Kuteva, 2007] B. Heine and T. Kuteva. *The Genesis of Grammar*. Oxford: Oxford University Press, 2007.
- [Hockett, 1960] C. F. Hockett. The origin of speech. *Scientific American* 203(3), 89–96, 1960.
- [Hopper and Traugott, 1993] P. J. Hopper and E. C. Traugott. *Grammaticalization*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press, 1993.



- [Hurford, 2003a] J. R. Hurford. The neural basis of predicate-argument structure. *Behavioral and Brain Sciences* 26(3), 261–283, 2003.
- [Hurford, 2003b] J. R. Hurford. Ventral/dorsal, predicate/argument: the transformation from perception to meaning. *Behavioral and Brain Sciences* 26(3), 301–311, 2003.
- [Hurford, 2011] J. R. Hurford. *The Origins of Grammar*. Oxford: Oxford University Press, 2011.
- [Katz, 1963] P. A. Katz. Effects of labels on children's perception and discrimination learning. *Journal of Experimental Psychology* 66, 423–428, 1963.
- [Kay et al., 1998] R. F. Kay, M. Cartmill and M. Balow. The hypoglossal canal and the origin of human vocal behavior. *Proceedings of the National Academy of Sciences of the U.S.A.*, 95(9), 5417–5419, 1998.
- [Kojima and Kiritani, 1989] S. Kojima and S. Kiritani. Vocal-auditory functions in the chimpanzee: Vowel perception. *International Journal of Primatology* 10(3), 199–213, 1989.
- [Kojima et al., 1989] S. Kojima, I. F. Tatsumi, S. Kiritani and H. Hirose. Vocal-auditory functions of the chimpanzee: consonant perception. *Human Evolution* 4(5) 403–416, 1989.
- [Leavens, 2004] D. A. Leavens. Manual deixis in apes and humans. *Interaction Studies* 5(3), 387–408, 2004.
- [Leavens and Hopkins, 1998] D. A. Leavens and W. D. Hopkins. Intentional communication by chimpanzees: a cross-sectional study of the use of referential gestures. *Developmental Psychology* 34, 813–822, 1998.
- [Levine and Meurers, 2006] R. D. Levine and W. D. Meurers. Head-Driven Phrase Structure Grammar. In E. K. Brown (Ed.), *Encyclopedia of Language and Linguistics, Second Edition*, pp. 237–252. Oxford: Elsevier, 2006.
- [Lieberman and Mattingly, 1989] A. M. Liberman and I. G. Mattingly. A specialization for speech perception. *Science* 243, 489–494, 1989.
- [Lieberman, 1984] P. Lieberman. *The Biology and Evolution of Language*. Cambridge, MA: Harvard University Press, 1984.
- [MacLarnon and Hewitt, 1999] A. MacLarnon and G. Hewitt. The evolution of human speech: the role of enhanced breathing control. *American Journal of Physical Anthropology* 109, 341–363, 1999.
- [Menzel, 2005] C. Menzel. Progress in the study of chimpanzee recall and episodic memory. In H. S. Terrace and J. Metcalfe (eds.), *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*, pp. 188–224. Oxford: Oxford University Press, 2005.
- [Ogden and Richards, 1923] C. K. Ogden and I. A. Richards. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. New York, Harcourt, Brace & World, Inc, 1923.
- [Payne and McVay, 1971] R. S. Payne and S. McVay. Songs of humpback whales. *Science* 173, 587–597, 1971.
- [Pepperberg, 2000] I. M. Pepperberg. *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots*. Cambridge, MA: Harvard University Press, 2000.
- [Pinker and Bloom, 1990] S. Pinker and P. Bloom. Natural language and natural selection. *Behavioral and Brain Sciences*, 13, 707–784, 1990.
- [Pinker and Jackendoff, 2005] S. Pinker and R. Jackendoff. The faculty of language: What's special about it? *Cognition* 95(2), 201–236, 2005.
- [Pollard and Sag, 1987] C. J. Pollard and I. A. Sag. *Information-Based Syntax and Semantics, Volume 1, Fundamentals*. Stanford, CA: Center for the Study of Language and Information (CSLI), 1987.
- [Pollard and Sag, 1994] C. J. Pollard and I. A. Sag. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press, 1994.
- [Ruhlen, 1994] M. Ruhlen *The Origin of Language: Tracing the Evolution of the Mother Tongue*. New York: John Wiley & Sons, 1994.
- [Saussure, 1916] F. de Saussure. *Cours de Linguistique Générale*. Paris: Payot, 1916.
- [Savage-Rumbaugh, 1986] E. S. Savage-Rumbaugh. *Ape Language: From Conditioned Response to Symbol*. New York: Columbia University Press, 1986.
- [Savage-Rumbaugh, 1990] E. S. Savage-Rumbaugh. Language acquisition in a nonhuman species: Implications for the innateness debate. *Developmental Psychobiology* 23(7), 599–620, 1990.

- [Savage-Rumbaugh, 1999] E. S. Savage-Rumbaugh. Ape language: Between a rock and a hard place. In B. J. King (Ed.), *The Origins of Language: What Nonhuman Primates can Tell us*, pp. 115-188. Santa Fe, NM: School of American Research Press, 1999.
- [Savage-Rumbaugh *et al.*, 1993] E. S. Savage-Rumbaugh, J. Murphy, R. A. Sevcik, K. E. Brakke, S. L. Williams, D. M. Rumbaugh, and E. Bates. *Language Comprehension in Ape and Child*. Society for Research in Child Development. (*Monographs of the Society for Research in Child Development*, Vol. 58, No. 3/4), 1993.
- [Schwartz *et al.*, 2005] B. L. Schwartz, M. L. Hoffman, and S. Evans. Episodic-like memory in a gorilla: a review and new findings. *Learning and Motivation* 36, 226-244, 2005.
- [Schwartz *et al.*, 2004] B. L. Schwartz, C. A. Meissner, M. L. Hoffman, S. Evans, and L. D. Frazier (2004). Event memory and misinformation effects in a gorilla (*Gorilla gorilla gorilla*). *Animal Cognition* 7, 93-100, 2004.
- [Seyfarth and Cheney, 1982] R. M. Seyfarth and D. L. Cheney. How monkeys see the world: A review of recent research on East African vervet monkeys. In C. T. Snowdon, C. H. Brown, and M. R. Petersen (Eds.), *Primate Communication*, pp. 239-252. Cambridge: Cambridge University Press, 1982.
- [Smith *et al.*, 1995] J. D. Smith, J. Schull, J. Strote, K. McGee, R. Egnor, and L. Erb. The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology* 124, 391-408, 1995.
- [Smith *et al.*, 1997] J. D. Smith, W. E. Shields, and D. A. Washburn. The uncertain response in humans and animals. *Cognition* 62, 75-97, 1997.
- [Smith *et al.*, 2003a] J. D. Smith, W. E. Shields, and D. A. Washburn. The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences* 26, 317-339, 2003.
- [Smith *et al.*, 2003b] J. D. Smith, W. E. Shields, and D. A. Washburn. Inaugurating a new area of comparative cognition research in an immediate moment of difficulty or uncertainty. *Behavioral and Brain Sciences* 26, 358-369, 2003.
- [Smith and Washburn, 2005] J. D. Smith and D. A. Washburn. Uncertainty monitoring and metacognition in animals. *Current Directions in Psychological Science* 14(1), 19-24, 2005.
- [Sperber and Wilson, 1986] D. Sperber and D. Wilson. *Relevance: Communication and Cognition*. Oxford: Blackwell, 1986.
- [Tincoff *et al.*, 2005] R. Tincoff, M. D. Hauser, F. Tsao, G. Spaepen, F. Ramus, and J. Mehler. The role of speech rhythm in language discrimination: Further tests with a non-human primate. *Developmental Science* 8(1), 26-35, 2005.
- [Todt and Hultsch, 1996] D. Todt and H. Hultsch. Acquisition and performance of song repertoires: Ways of coping with diversity and versatility. In D. E. Kroodsma and E. H. Miller (Eds.), *Ecology and Evolution of Acoustic Communication in Birds*, pp. 79-96. Ithaca, NY: Comstock Publishing Associates, 1996.
- [Todt and Hultsch, 1998] D. Todt and H. Hultsch. How songbirds deal with large amounts of serial information: Retrieval rules suggest a hierarchical song memory. *Biological Cybernetics* 79, 487-500, 1998.
- [Trivers, 1971] R. L. Trivers. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46(4), 35-57, 1971.
- [Watson *et al.*, 2001] J. Watson, G. Gergely, V. Csanyi, J. Topal, M. Gacsi, and Z. Sarkozi. Distinguishing logic from association in the solution of an invisible displacement task by children (*Homo sapiens*) and dogs (*Canis familiaris*): using negation of disjunction. *Journal of Comparative Psychology* 115, 219-226, 2001.
- [Xu, 2002] F. Xu. The role of language in acquiring object kind concepts in infancy. *Cognition* 85, 223-250, 2002.
- [Zoloth *et al.*, 1979] S. R. Zoloth, R. Petersen, M. M. D. Beecher, S. Green, P. Marler, D. B. Moody, and W. Stebbins. Species-specific perceptual processing of vocal sounds by monkeys. *Science* 204, 870-873, 1979.
- [Zuidema and de Boer, 2010] W. Zuidema and B. de Boer. The evolution of combinatorial phonology. *Journal of Phonetics*, 37(2), 125-144, 2010.