2
Evolutionary linguistics
How Language and languages got to be the way they are

James R. Hurford

2.1 Introduction

‘Nothing in biology makes sense except in the light of evolution’ is a famous dictum of the biologist Theodosius Dobzhansky. To ‘make sense’ of a phenomenon is to (begin to) explain it. A high goal of linguistics has traditionally been to explain language (an ambiguous term, as we will see immediately), rather than merely to describe it. Given this high goal, it is vital to be clear about the thing or things to be explained, the explanandum or explananda of our subject. So linguists must distinguish between two senses of ‘language’. One is the biologically given human faculty of language. This faculty is what constitutes the language-readiness of a normal newborn infant, prepared to absorb the torrent of experience assailing it from the buzz of conversation around it, and the associated activities, and after a few years to become a competent participant in this buzz and these activities. In this chapter, I will refer to this biologically given language faculty with a capitalized mass noun, ‘Language’. The other sense of ‘language’ is any of the entities that historico-cultural processes have delivered to the world, that is, particular individual languages, such as Albanian, Dyirbal, Hmong, Navaho and Zulu. I will refer to them generically using a count noun, as in ‘many languages’ or ‘this language’. Each language is a body of complex conventional behaviour (highly articulated spoken and written buzz and associated activities) in a more or less clearly defined community. The conventionality of the behaviour implies that each participant has mentally internalized a collection of normal patterns much like that also learned by his fellows in the community.

Clearly, different kinds of explanation are needed for the biological faculty, Language, and the historico-cultural entities, languages. Just as clearly, the thumb-nail definitions of them I have started with make some tricky presuppositions and need further amplification. And to complicate things further, it turns out that the biological and historico-cultural processes are to a degree intertwined, in processes of gene-culture co-evolution. The necessary amplifications and added complexities will be touched on as we proceed.

As this chapter is largely targeted at linguists, it will be useful to carve linguistics near to its commonly understood natural joints, so without too much revision of the usually assumed boundaries, it will be possible to deal with, for example, evolutionary semantics, evolutionary
pragmatics and evolutionary phonetics. Equally, but calling on different explanatory mechanisms, and discussed in less detail here, one may distinguish evolutionary phonology and evolutionary morphosyntax. Any revision of boundaries is motivated by the distinction between different modes of explanation, biological and historico-cultural. I will take as biologically given to each normal newborn human the semantic, pragmatic and phonetic prerequisites of the Language faculty. Filling that out a bit, and still roughly speaking, the semantic prerequisites for Language include a faculty for concept formation, mental representation of propositions about the world, and computation of inference, the pragmatic prerequisites include a disposition to cooperative engagement with others and an ability to guess their intentions rather accurately, and the phonetic prerequisites include coordination of the acoustic patterns of speech with the motor control necessary for producing and decoding articulate sounds. All this, and more, has to be in place in a baby if it is to become a fluent speaker of its group’s language. And all this had to be in place, at least in some rudimentary form, in the earliest hominin ancestors of *Homo sapiens*, otherwise Language as we know it would not have evolved. We will see small evolutionary seeds of these prerequisites for Language in the behaviour of non-human animals, more or less directly genetically related to humans.

By contrast, the complex morphosyntactic and phonological structures of individual languages are the products of centuries, in most cases many millennia, of culturally transmitted, and culturally enhanced, patterns of behaviour. A child experiences the behaviour of older people already regularly using and productively combining established words and constructions, and begins to use and productively combine the same words and constructions, perhaps with some tiny idiosyncratic innovations, which may or may not survive to be passed on in future generations. This conceives of evolutionary morphosyntax and evolutionary phonology as largely matters of cultural evolution, successive waves of learning and cumulating small innovations. There can be no doubt that, in the long view of prehistory at least back to the emergence of our species, the syntactic and phonological complexities that we see in languages today emerged out of something simpler. At the end of this chapter, I will very briefly argue that we can speculate rather confidently about what such simpler proto-forms of languages were like.

In this introductory delineation of the different mechanisms called on to explain Language and languages, we immediately see an instance of the phenomena studied by linguistics beginning to make sense in the light of evolution, as they would not if evolution were not considered. The rest of this chapter will discuss each of the branches of linguistics in suitable, biological or cultural, evolutionary terms.

Niko Tinbergen suggested four questions that need to be answered in seeking explanations for any biological phenomenon. These questions can be posed of the human faculty of Language, in the following ways.

### 2.1.1 Mechanism

How does the organism work? What are its parts and how do they interact with each other? Specifying the mechanism of languages is an interdisciplinary enterprise, in which linguists, psycholinguists and neuroscientists must collaborate. Linguists describe languages in terms specially developed, often over many centuries, for the specific task of describing languages, terms such as ‘noun’, ‘verb’, ‘clause’, ‘passive’, etc. Such terms are not part of the vocabulary of psychology, neuroscience or biology. Linguists’ descriptions, quite abstracted from neural mechanisms, are nevertheless intended to be ‘psychologically real’ in the sense that
neural correlates will presumably eventually be found for the parts of the linguist’s descriptions. Some subtlety and skill will no doubt be involved in mapping the linguist’s abstractions onto neural processes. Certainly, naive localization of ‘addresses’ in the brain where, for example, nouns ‘can be found’, is to be avoided. The more plausibly a neurological interpretation of the linguist’s descriptions can be defined, the greater the psychological justification of the linguist’s original descriptive apparatus. It is assumed that the mechanisms underlying all individual languages are significantly similar. When a person uses a language, mechanisms of both storage and computation are involved. Words and constructions are summoned up from memory, and in production combined by the brain into more or less complex structures. In language recognition, the same memory store is used, and the brain analyses received complex input, mapping it onto understandings of what was said, or even of what the speaker intended. All such specifically linguistic mechanisms collaborate in the brain with more language-peripheral processes, such as understanding and management of the social and physical context.

2.1.2 Function

What does the organism do? In the case of languages, they serve both for public communication between people and to some extent for the facilitation of private thought. It needs to be shown how the mechanism, the parts of a language and their workings, are applied to carry out these functions. The study of how precisely language facilitates private thought, relying on some independent characterization of what thought itself consists of, is challenging, and as yet little developed. It is the domain of linguists, psycholinguists and neuroscientists, with the possibility of some useful input from philosophers. The study of how languages serve communication is rather less challenging, just because communication is public. Philosophy, linguistics and even folk theories converge on ideas relevant to the communicative functions of language, ideas such as reference to objects and events, speech acts, and information structure. Parts of language (e.g. common nouns, verbs and adjectives) denote classes of objects and events, other parts (e.g. proper nouns and definite descriptions) pick out specific individuals, other parts (e.g. function words) serve as traffic signals to a hearer helping the parsing of an utterance, other parts (e.g. interrogative markers) signal what speech act is being carried out.

2.1.3 Ontogeny

How does the organism develop in an individual? In biology, this is the province of embryology, charting the route from a single cell to an adult organism. Developmental linguistics, studying children’s acquisition of languages, can hardly start so far back. There is no linguistic analogue of the single cell. Instead, developmental linguistics has to make do with the first behavioural signs of language in newborns. Even in neonates, there is evidence of the influence of experience in the womb, with babies apparently sensitive to the different types of rhythm found in languages. The different subsystems of languages, their phonology, syntax and lexicon, have different sensitive periods for their development. Accurate native-like pronunciation is extremely difficult to achieve after about ten years of age, much syntax can be acquired after that age, though with more effort than earlier, and the acquisition of new vocabulary can continue into late life. In all cases, complex developments are built upon simpler foundations. Children start with one-word utterances, proceed to two-word utterances, and then move on quite fast. Likewise with syllable structure,
consonant–vowel (CV) syllables being the basis upon which other more complex syllables are built. The progression from simple units to more complex ones is a necessity of the learning process. Complex cases cannot be learned without the prior learning of simpler building blocks.

2.1.4 Phylogeny

How has the organism evolved in its species? This, obviously, is the domain of any putative evolutionary linguistics. Evolutionary linguistics is less well developed than other aspects of the subject, partly for the very good reason that solid evidence about the evolution of the Language faculty is hard to come by. Language leaves no fossils, in the sense of remnants of behaviour that can be dug up. Tinbergen was not concerned with any phylogeny other than the biological kind, the evolution of a species, as he was a biologist not concerned with human culture. Cultural transmission of behaviour in a group is far, far more developed in humans than in any other species. And human languages, in particular, are unique in life on Earth as complex culturally evolved artefacts. It seems reasonable to stretch Tinbergen’s heading of Phylogeny to include the cultural evolution of languages as well as the biological evolution of the Language faculty. In speculating about the early evolution of syntactic and phonological systems, it is reasonable to give some attention to Haeckel’s maxim that ‘ontogeny recapitulates phylogeny’. That is, what is seen early in child language development is also likely to have occurred early in the emergence of the first languages in our species, probably around 150,000 years ago. There is no logical necessity for ontogeny to recapitulate phylogeny, and sometimes it does not. But the principle of simpler beginnings being the foundation for more complex structures also applies in phylogeny. In the case of language, modern children have the advantage of learning an existing system. For the phylogenetic story we wish to tell, the focus is necessarily on innovations not drawn from any existing system. Cases where modern children naturally come up with expressions that are not found in the ambient adult language are likely to be especially valuable as clues to the innovative possibilities that have driven the cultural evolution of languages.

2.2 Evolutionary semantics

Semantics deals with what Halliday (1973) usefully called ‘ideational’ meaning. Among other things, this concerns the denotations of predicating words, the referents of referring expressions, the combinations of these into meaningful propositions, and relations of valid inference among propositions. The evolutionary question is: where did this stuff come from? Naturally assuming that these are not metaphysical or Platonic entities, we take a mentalist approach to the question. That is, we suppose that relations between words in a language and things in the external world are mediated by mental representations in the heads of language users. A mentalist approach to predicates, reference and propositions is outlined in more detail just below.

Predicate words, like person, red, run, hit, give and above are linked to entrenched concepts in users’ minds, and these concepts in turn relate to entities, properties and relations in the outside world. In describing something as red, for example, a speaker judges that the referent matches a certain range of the colour spectrum categorically defined in his mind, and distinguished there from other regions of the spectrum, denoted by other colour terms. Likewise, one carries a mental distinction between what counts as a person and, say, a rock or a tree. To a large extent, such mental categories are formed and kept distinct by
nonlinguistic, perceptual experience of the world, although what words are appropriate to label them is, of course, a matter of linguistic experience. Imagine you are watching birds out of your window. You see various different repeating types, and perhaps note their different habits. You can do this without knowing what the different bird types are called. Likewise with flowers, trees, spicy tastes, smells, various sounds of nature, many of which you can recognize and keep distinct, without knowing their names. Thus, we assume that there exist mental acts of categorical judgement independent of any sentences in a language that might express predication. It is reasonable to call such judgements acts of ‘nonlinguistic predication’. The qualification should be enough to satisfy any who insist on keeping predication pure and simple as a solely linguistic phenomenon. Mentally categorizing individual objects with their properties is relatively simple. Judging relationships between objects is more complex, obviously, because it involves more objects. Judging two objects to be of the same, or different, category involves first making individual judgements about each separate object, then mentally comparing the judgements arrived at. Another more abstract predication is that expressed, for example, by *Red is a colour*. This is more abstract because redness is a first-order property (predicate) of objects, such as ripe cherries and cricket balls, and being a colour is a second-order property (predicate) applicable to redness, along with blueness, greenness, etc. It is not clear whether this latter type of abstract predication is available to us independently of language. But, given language, it certainly is.

The most basic referring expressions, deictic terms such as *this* and *that*, in typical uses, are the linguistic expressions of mental indices provided by the attention system. These indices are temporarily linked to whatever object or event in the world an individual happens to be attending to at the time of use. The attention mentioned can be effected through any of the senses, but typically through vision and audition, which provide the possibility of joint attention by different individuals. Thus, if I say ‘What’s that?’, I have some visual or auditory stimulus temporarily in mind, an uncategorized sight or a sound, referred to by the pronoun *that*, and I assume that my interlocutor can figure out from her own ongoing experience of the world which sight or sound I am asking about. The next time someone asks ‘What’s that?’, it will almost certainly be about a different sight or sound, to which the speaker has attached a temporary mental index, again expressed in use as *that*. In the terms adopted here, acts of paying attention to things are mental acts of reference. Attention to something need not be accompanied by any linguistic act.

Less basic, non-deictic referring expressions, such as *The vicar’s cat* or *Kevin* pick out individuals in the world by virtue of the user having a mental representation of the individual concerned as unique in its context of use. Each time someone uses *the vicar’s cat* they may well be talking of a different animal or even a different vicar. But if communication is successful, the mental representations that users have of these referents are clear and well separated in their minds from other kinds of things, such as dogs and kangaroos. My mental representation of a particular animal, whom you or I might choose on some occasion to call *the vicar’s cat*, and on another occasion *Kevin*, is distinct in my mind from my representation of other entities, whom people might call *the professor’s dog* or *Amelia*. A person knows a whole lot of entities, their properties and behaviour, and this knowledge is largely, though not entirely, independent on language.

In logic, propositions are composed from two types of term, a predicaing term and one or more referring terms. (A seldom discussed exception is the simplest type of proposition, with no referring terms, such as propositions about the weather, as expressed, for example, in *It’s raining*.) Thus, for instance, *That is a cat* and *Kevin is furry* both express propositions. Propositions are sometimes identified as facts about a world, existing somehow independent
of knowers and language. Taking a mentalistic view, a proposition is a mental combination of a category and a mental index. The index is the mental referent, the category the mental predicate. We can know propositions, without ever having expressed them in language, even subvocally. For instance, I know that a certain sort of grey caterpillar habitually munches a certain kind of plant in my garden, but I do not know the names of either caterpillar or plant. Oops, now I have just expressed that proposition linguistically, albeit with circumlocution, but before I did, I knew the proposition independently of language, an observed fact about those caterpillars and those plants. There surely are some facts (propositions) about the world that are either unobservable or ineffable, or both. We will never know everything, let alone be able to say it. Obviously, from the point of view of language, we are only interested in such propositions as can in principle be expressed in language.

Not only do we hold propositions in our heads, we can also manipulate them in chains of inference. If you know that Kevin is a cat, you know he is a quadrupedal mammal. More complexly, if you know Kevin is tougher than Amelia, my cat, and that Amelia is tougher than Sherman, my creampuff dog, then you also know that Kevin is tougher than Sherman. This shows that you can do transitive inference, among many other kinds of inferential calculations involving propositions.

Now, to repeat the evolutionary question: where did this stuff come from? Having taken a mentalist viewpoint, we cannot shrug the question off by assuming that, ever since the Big Bang, referents, predicates and propositions always existed, timelessly, even though there were not always brains adequate to entertain them. Their mental correlates have only existed for as long as there have been brains powerful enough to house and process the corresponding neural representations. At least, the existence of brains, or central nervous systems of some kind, is a convenient starting point here, though it is debatable. We cannot yet push our evolutionary question further back than the emergence of brains, say to protozoa. And even the emergence of brains does not necessarily immediately give rise to the type of fully-fledged semantic mental representations that are a basis for language.

The availability of mental indices for variable and versatile attachment to objects of attention is a prerequisite for dealing practically with the world, in any somewhat advanced animal. An animal manipulating, say, a stick, to fish termites out of a mound is attending to the stick and the hole in the mound. At other times, say when grooming another ape, the attentional indices may be linked to particular patches of skin and lice therein. A chimp spotting a louse in its friend’s fur makes a judgement about the thing it has seen – not a freckle, not a hair, but a louse, worth catching. Such instances are fleeting, but are the basis for more permanent mental representations. A social animal that recognizes particular fellow troop members and behaves differentially and systematically toward them, e.g. by bullying them, or avoiding them, or grooming them, has formed somewhat permanent mental indices of particular individuals. In truth as well as proverb, elephants never forget. What this means, at the least, is that elephants can preserve, over many years, representations of particular individuals, who have behaved either well or badly toward them, be they humans or other elephants. Long-term memory for individual objects, linked to their significant properties, is uncontroversial among mammals and birds. Non-human animals (apart from honey-bees) do not communicate to each other about things distant in space and time, but they do clearly have memories about things distant in space and time. Starting to communicate about what they remember and know is another matter.

A more complex type of memory than memory for objects and their properties, a memory for specific events involving several individuals, is more debatable. This is known as
episodic memory, and was once claimed to be unique to humans. It is the kind of memory impaired in some amnesics. Clearly, humans have a far greater capacity than other animals for remembering and recalling events from the past. Experiments with a wide range of animals, including apes, and some birds (scrub jays in particular) have shown that they have the ability to remember for a short time events they have observed or in which they have participated. A chimpanzee called Panzee was able to signal to a trainer the location of a kiwi fruit that another trainer had hidden the day before. Scrub jays have been shown to remember where they cached perishable food, and only to retrieve it if not too much time has elapsed since it was cached. By contrast they remember where non-perishable food was cached and go back to it after longer periods. As the researchers put it, the scrub jays remember the what (perishable/non-perishable food), the where (the caching spots), and the when (how long ago) of the caching events. Of course, human capacity for remembering events is much more far-reaching, in time, space and the kind of events recalled. But the existence of some kind of episodic memory in other animals, however simple, shows that this aspect of the mental representation of facts about the world is quite ancient, possibly dating from the emergence of birds and mammals, or possibly having evolved separately several times since.

Many non-human animals clearly engage in some degree of advance planning of their actions. This can involve representation of places, movements, objects and possibly planned complex events. Animals do not confuse planning of future actions with memories of past actions, although the actions themselves may be of the same type. Thus, an implicit understanding of the difference between past and future is attributable to non-human animals. Surely, this should be no surprise.

As for somewhat complex predications, those involving sameness or difference, a very wide range of animals has been shown to be capable of making same/different judgements. Some animals, for example one California sea lion, remembered a same/different task it had been trained on ten years earlier. Among symbol-trained animals, Alex the African grey parrot was capable of the kind of abstract judgement involved in Red is a colour. This is clear because experimenters would show Alex an object, and ask What colour is this? and he would reply, as appropriate, red or blue, etc. Note that the question was not the first-order question Is this red?, but relied on Alex making the mental connection between the perceived redness of some object, and the higher-level ‘fact’ that red is a colour. Now, Alex had been taught a set of language-like symbols, spoken words in his case. It has not yet been shown that any non-human animal is capable of this kind of second-order predication independent of having learned some overt symbols for the first-order predicates, e.g. red, blue.

Non-human animals are capable of a degree of computation involving the propositions they know. Social monkeys, such as baboons, know the pecking order of their troop in detail. They know, for any given pair of animals, which is dominant to which. It is likely that this knowledge is partly learned by direct experience of acts of aggression and submission, but also partly by transitive inference. The baboon thinks something like ’X submitted to Y, and Y submitted to Z, so X would submit to Z’, even though it has not observed any interaction between X and Z. This capacity has been attributed to chickens as well.

An interesting experimental task involves reversal learning. An animal is first taught specific connections between two stimuli and two responses. Of course, many animals can be trained in such tasks, though their speed of learning differs considerably. Then the connection between stimuli and responses is deliberately reversed. So whereas previously it was correct to press bar X in response to stimulus A, and bar Y for stimulus B, now the new required connections are X for B and Y for A. Animals adapt to this reversal with differing
degrees of aplomb. There is a significant difference between apes and (most) monkeys. For monkeys, the better they have learned the first set of connections, the more trouble they have in unlearning it and picking up the new connections. For apes, by contrast, the better they have learned the first set of connections, the quicker they adapt to the new task. A plausible explanation is that the apes, but not the monkeys, have formed a mental rule, and are able to apply an oppositeness operator to what they had first learned. Informally, ‘Ah, I see, it’s the opposite of what I was taught before’. Apes are genetically closer to humans than monkeys, and this difference in their learning abilities probably reflects an evolutionary development in our lineage toward more abstract mental computation.

In sum, much of the basis of human semantics is already present, though to a simpler degree in non-human animals. Non-human animals have somewhat rich mental representations of the world. They just do not communicate these to others. Communication with others is a matter for pragmatics, not evolutionary semantics, as here defined. The first steps in the evolution of our semantic abilities, taken long before the emergence of our own species, involved developing the mental capacity to store and compute with simple propositions, even somewhat abstract ones.

Of course, semantics to linguists is defined as the study of meaning in languages. So the linguist may feel cheated at this point in that I have traced the origins of semantics to a stage before language. I have identified meanings that get conveyed by words and sentences as private mental entities pre-existing the public application of labels to them for purposes of communication. The process of developing a communal shared vocabulary to describe communally shared types of experience is straightforward, and has been simulated in many computer models. This social process requires, of course, the prior existence of individuals with the apparatus for representing meanings, as outlined above, plus some cooperative disposition to communicate meanings.

In a significant minority of cases, meanings of words are the product of, not the precondition for, the social naming process. These include cases of technically defined words, such as dodecahedron, hundred and thousand. The exact concept of a hundred is not accessible without language, specific words with simpler meanings, such as two, hands and ten. We are explicitly taught that a hundred is ten tens. Once the concept of a hundred is established, a further linguistic definition of thousand is possible. So the evolution of some aspects of the meaning in languages is a matter of cultural evolution, far less ancient than the biological evolution of the basic capacities I have discussed above, and postdates the emergence of some language.

### 2.3 Evolutionary pragmatics

Halliday (1973) distinguished ideational meaning from interpersonal meaning and rhetorical meaning. The latter two belong in what is now gathered together under pragmatics, whose central topics include the theories of speech acts, implicature and information structure. Speech act theory, starting from J.L. Austin’s work (1975 [1962]), emphasizes that when we speak we not only describe aspects of the world, but we also attempt to do things to the hearer, to influence her mind or his behaviour in some way. The theory of implicature, due originally to Paul Grice (1975), explains how apparently illogical assertions can nevertheless usefully convey meaning, because hearers are able to infer conclusions based on a premise of cooperativeness. The working out of implicatures, like the interpretation of indirect speech acts, involves a degree of mind-reading by interlocutors, the making of assumptions about what each party is likely to know and to want to know. Information structure is the
way propositional information is presented. The same proposition can be expressed in a
great variety of different ways, tuned to the assumed shared knowledge and conversational
goals of the hearer. How ancient are our abilities to handle and interpret speech acts, to
manipulate and interpret implicatures, and to structure the way information is presented? As
in the case of semantics, the seeds of the necessary human abilities in two of these areas,
speech acts and implicature, are not unique to us, but are present in many non-human
animals, especially those closely related to us. Thus the evolution of much of pragmatics has
ancient roots, pre-dating the emergence of *Homo sapiens*. The development of information
structure is more recent, co-evolving with the rise of somewhat complex syntax in languages.

Many animals communicate, even those very distantly related to us, such as insects. With
a tiny number of exceptions, all such communication is dyadic, in the sense of only involving
two parties, the sender and the receiver of the communication. Most animal communication
is not triadic, in the sense of involving some third entity in the outside world being
communicated about. Vervet monkey alarm calls, along with the alarm calls of other
monkeys and many species of bird, are triadic, in that they are triggered by entities outside
the sender and receiver(s), namely the predator in question. Some primates also give food
calls, indicating the presence of various types of food. None of these triadic calls are, as far
as we know at present, learned. Rather, they are almost entirely innate, in the sense of being
made by the animals without any prior learning experience. The slight reservation here is
due to the fact that vervet babies do refine the category of predator to which they give alarm
calls during their early life. The honey-bee waggle dance, also triadic, in indicating the
whereabouts of food, is entirely innate, too. Humans are the only (non-captive) species who
learn triadic signals. And we learn tens of thousands of them. Captive animals, from
chimpanzees and dogs to parrots, can learn small vocabularies of referring words, but this
ability is never exploited in the wild. Part of the story of the evolution of *Language* is the
emergence of this impressive human learning ability.

In dyadic communication, an animal does something to another, be it threatening,
submitting, expressing solidarity, keeping in touch with the group or inviting mating, for
example. These acts do not describe or refer to anything outside the sender or receiver. Thus
they are analogous to bare illocutionary expressions in human languages, such as *Hello,\nWhoa!, Ugh!* and *Bye*. All languages have expressions of this bare illocutionary dyadic kind.
Indeed, in any language, all utterances addressed to a hearer are made with the intention, on
the part of the speaker, to affect the hearer in some way, be it to warn him, to congratulate
her, to advise him, to cheer her up, or to insult him. Why else would anyone speak to
another, if not to achieve some effect? (Occasional solipsistic use of language, as when one
utters to oneself, clearly does not deliberately affect another person, but such uses are
plausibly derivative of public utterances.) This pervasive doing-things-to-each-other aspect
of using a language builds upon similar communicative acts by our remote non-human
ancestors. Since human communication took on its mainly triadic function, with the vast
majority of utterances involving reference to one or more entities outside the speaker and
hearer, a greater range of illocutionary acts have become possible. Only with complex
language, for example, can one make a promise, apologize explicitly for a misdeed, formally
name a ship, or perform a legal marriage.

The direct illocution of an utterance is coded in its syntactic structure. Thus, English
sentences in interrogative form, with inverted subject and auxiliary, have as their direct
illocution the posing of a question, soliciting information. Often, however, the indirect
illocution in uttering such a sentence may be other, such as to request an action (e.g. *Can you
pass the salt?*) or point out the obviousness of some fact (e.g. *Is the Pope Catholic?*), and
How are you?, merely used as a routine greeting, and so forth. Some such cases are conventional and idiomatic, but others require some mind-reading on the part of the hearer. More generally, human language users are capable of drawing indirect conclusions from what is said, where these conclusions are not explicitly coded in what was said. This is implicature. If a mother asks her child Have you done your homework and tidied your room? and the child answers only Well, I’ve done my homework, the mother can reasonably conclude that the child has not tidied his room. Coming to this conclusion involves some rather complex reasoning, including making assumptions about what the child might have said if he had indeed tidied his room. The mother, like other language users, depends on knowing what is in the mind of another speaker, a kind of (non-telepathic) mind-reading.

Non-human animals can to some degree read the minds of others. ‘Mind-reading’ is not to be taken in any mystical sense. It involves being able to assess the intentions of others and predict their likely behaviour. In well-controlled experiments, chimpanzees have been shown to judge accurately where a dominant competitor is likely to go. In other experiments, chimps have been shown able to distinguish between when a human interactor is teasing them or merely fumbling, when the two actions are minimally distinct visually. Mind-reading in this sense involves a certain kind of self-knowledge. Thus chimpanzees are able to infer behaviour in others just where their own behaviour would be similar. For instance, chimpanzees are generally competitive and less cooperative than humans, and chimps are much better at reading another’s competitive intentions than they are at anticipating a cooperative action. None of this, of course, involves language of any kind, but such mind-reading is at the heart of any account of how humans interpret the utterances of others, whether through calculating implicatures or getting to the indirect illocutions of speech acts. Human capacities in these domains has built upon pre-existing capacities in pre-human ancestors. Doubtless, we are better and subtler at it than our ancestors were, let alone our chimpanzee cousins.

Humans are naturally more disposed to cooperate with fellow members of their group than other species. Unless cruelly repressed, children show an instinctive drive to join in communicative exchanges. At first, not having learned the complex wherewithal, i.e. the necessary vocabulary and grammar, they resort to pointing. A remarkable difference between human infants and chimpanzees is that human toddlers spontaneously point to things merely, apparently, to share joint attention – ‘Look!’ In the wild, chimpanzees never point at things, and in captivity, they only point for the purpose of begging something from a human: even in captivity, chimpanzees do not point for fellow chimpanzees. This human characteristic has been labelled ‘shared intentionality’.

The rhetorical manipulation of information in languages is manifested in the availability of a wide range of sentence structures, intonation patterns and lexical choices. There are literally dozens of different ways of presenting the proposition that a lorry crushed a car, depending on what has just been said in the conversation and where a speaker wants the conversation to go next. The syntactic devices used in English include passivization, topicalization, What-cleft sentences, It-cleft sentences, and combinations of all these. Examples are It was the lorry that crushed the car, As for the car, it was what was crushed by the lorry. Some such combinations are cumbersome, but English makes them available, and one can think of conversational contexts in which they would be appropriate. Children learn to control such complex cases late, pidgin languages lack such grammatical complexity, and it is reasonable to suppose that the growth of such devices was at least somewhat gradual, with the earliest languages of our species having few or none of them.
2.4 Evolutionary phonetics

Without concepts to express, and without the social motivation to express them, Language would not have got started. The speech medium is not absolutely crucial for the expression of complex meanings, as well-developed sign languages exist, with a similar expressive range to spoken languages. It is likely that the very earliest external forms of human language incorporated manual gestures, and the move to almost total dominance of speech followed over the millennia. Chimpanzees have great manual dexterity, under voluntary cortical control, and have nothing of the delicate vocal articulatory control that humans show in speech. The brain areas, predominantly left hemisphere, involved in complex language are close to the areas managing gestures with the dominant right hand. Gestures can be much more easily used in iconic, easily interpretable, ways to communicate actions and the shapes of objects, whereas the iconicity of speech is limited to a few natural sounds, mainly a small range of animal calls, e.g. *miau*, *moo*. Thus if modern languages were universally signed, rather than spoken, there would be little to explain about the medium itself. Given concepts to express, and the will to express them, manually gestured communication would emerge quite naturally from the dexterity of our primate ancestors, and their facility in deliberately using them. The primate visual system needs no special adaptation to perceive manual gestures.

Of course, speech has its advantages too. It can be used in the dark, with one’s back to an interlocutor, and around corners. Another important advantage of speech is its apparent uselessness for any other purpose. A waving gesture might be interpreted as not carrying any meaning, but just waving insects away. A touch of the fingers to the cheek could be fixing an itch. Rubbing the hands together could be for generating some warmth. For a full sign language to emerge, a division must be established between practical non-communicative movements and conventionally meaningful ones. This can happen, as the emergence of natural sign languages shows. But speech, being pretty much useless for any practical non-communicative purpose, is likely to be taken as expressing meaning from the start, given suitable expectation that the actor could well be trying to say something meaningful. Speech rather naturally signals its own signalhood.

The evolution of articulate speech from primate vocalizations is a much greater step than the evolution of manual gestures. Our nearest primate relatives can manage essentially nothing in the speech domain. But some brain basis for the later evolution of meaningful speech is found in the fact that monkeys and apes generally process the meaningful calls of conspecifics in their left hemispheres, rather than bilaterally, as with other sounds. Over the roughly six million years since humans diverged from chimpanzees, the human vocal tract, and our complex control over it has evolved significantly. The main changes identified in the literature are: (1) lowering of the larynx in humans; (2) development of breath control; and (3) development of fine articulatory control. In addition, there are some small differences between human and chimpanzee hearing that reflect a likely adaptation to the range of vocal sounds that humans began to make. We will briefly deal with each of these.

The human larynx, almost but not quite uniquely among mammals, is low down in the vocal tract, relatively far from the junction of the oral and nasal passages. Thus in most mammals, any vibration initiated by the larynx only has the mouth chamber, or less importantly the nasal passages, to resonate through. With the human larynx far down in the neck, vibrations are channelled through a more complex L-shaped tube, first the pharynx, in the throat behind the tongue, and then outward through the mouth. It is this double resonator tract that makes it possible to produce different vowel qualities. A high front [i] vowel is
produced with a widened pharynx and a narrowed mouth chamber, as the tongue body is pushed upward and forward. Contrastingly, a low back [a] vowel is produced with a narrowed pharynx and a widened oral cavity, the tongue being drawn downward and backward. In between these two extreme vowels, a range of other vowel qualities are available by different intermediate positions of the tongue and lower jaw. Rather prehensile lips also help to increase the acoustic variety that the human vocal tract is capable of; chimpanzees share prehensile lips with us, but the range of vowel-like qualities producible with only alteration of lip protrusion is limited. Over the last six million years the position of the adult human larynx has lowered from that typical of our closest relatives.

Although, as mentioned earlier, ontogeny does not necessarily recapitulate phylogeny, in the case of the position of the larynx, it does. The larynxes of human newborns are high up just behind the velum, where the oral and nasal passages meet, as in most adult mammals. This position allows the baby to breathe and suckle at the same time. In the first six months of life, the baby’s larynx descends to nearer its eventual adult position, in this case recapitulating phylogeny. The timing during human evolution of the larynx’s descent is hard to pin down, as soft tissue does not fossilize. But indirect evidence from the hyoid bones of Neanderthals suggests that this descent was in place in the common ancestor of humans and Neanderthals, dated tentatively to half a million years ago, and perhaps as much as a million years ago. The hyoid bone sits in the throat above the larynx, which it partially supports.

Breathing during speech is heavily imbalanced between in-breaths and out-breaths. While speaking, we breathe out in a highly controlled manner for about 90 per cent of the time, with quick pauses for inhalation. Fine control of breathing is mediated by the chest and stomach muscles. It is likely that this degree of control was made possible by bipedalism, which frees the upper torso from dependence on bracing the chest to take the impact of moving on the front limbs. Chimpanzee ‘laughter’ is a rhythmic see-sawing in-and-out breathing vocalization, with as much time on the in-breath as the out-breath. But human laughter, like human speech, is all on an out-breath, even though it has a similar rhythm to chimp ‘laughter’. The Australopithecines of four million years ago walked bipedally, so this facilitating relaxation of constraints paving the way for an aspect of speech was in place very early in our evolution. Studies of the diameter of the vertebral canal, through which the nerves to the chest muscles pass, in modern specimens and in fossils, confirm that Neanderthals could have had similar breath control to modern humans, in contrast to Homo erectus and Australopithecines. Human operation of the vocal articulators, the tongue, jaw and velum, in coordination with well-controlled vibration of the vocal cords, is very delicate. Some speculative studies have attempted to relate this to the size of the hypoglossal canals, the little holes in the base of the skull through which nerves serving the tongue and jaw pass. Here, it has not been possible to identify any significant differences between modern humans, hominin ancestors and related primates. The era when fine control of the vocal articulators emerged has yet to be determined.

Human hearing is in most respects very similar to chimpanzee hearing, but there are small differences, relevant for sensing nuances of vowel quality. The most salient difference is that in the region of about 34000 Hz, chimpanzee auditory acuity is significantly poorer than human acuity. This acoustic region is important for the recognition of high front vowels such as [i] and [e], whose second formants are typically near this frequency. At some point since the chimp–human split, we developed acuity in this frequency range, and this could well be an adaptation to the range of vowels the human vocal tract, with its lowered larynx, is capable of.
2.5 Evolutionary morphosyntax and phonology

The previous sections have dealt exclusively with factors delivered by biological evolution, facilitating the subsequent cultural evolution of complex languages. In this final section, I will deal, more briefly, with the cultural evolution of the complex grammatical and phonological patterns that occupy descriptive linguists.

For the cultural evolution of languages to begin and continue, certain biological capacities had to be in place and to grow. The complex syntax of any language requires that individual speakers have a powerful storage capacity for the large vocabulary of basic items it combines. And individual speakers also need the computing power to recursively combine words and constructions up to quite impressive (though of course not infinite) levels of embedding. These capacities for vast storage and complex computation, aspects of Language, are not learned but encoded somehow in the DNA determining the growth patterns of the modern species. It is likely that during the cultural evolution of complex languages from simple beginnings over the past 150,000 years, these biologically given capacities also expanded somewhat, under pressure from the growing practices of the groups using the languages concerned. Co-evolution of genes and culture is well attested in cases outside L/language, and in fact it can happen very fast. The evolution of lactose tolerance in pastoral groups over at most 20,000 years is a case in point. Over many centuries, languages grew more complex grammars, and correspondingly, due to the great practical advantages of language, facility for managing such complexity grew in human brains.

Languages are not all equally complex in their grammar. This is now widely accepted, contrary to a long-standing dogma in linguistics. Simpler grammar does not necessarily entail poorer powers of expression. Some complexities of grammar contribute little or nothing to actual expressive power, as telegraphic language (e.g. newspaper headlines) testifies. And there is always a flexible negotiation between how much information is carried explicitly in the coded norms of a language and how much can safely be left to pragmatic inference from context.

Complex grammar arises by grammaticalization. This much is a tautology. Grammaticalization is simply the historical process by which grammar grows. Here, we can break grammaticalization down into two separate but closely related processes. By one process, a distinction emerges between function words and content words, and the set of function words expands, with concomitant increase in the range of different grammatical constructions signalled by these function words. By another process, free words get shortened and phonologically bound to content words, giving rise to the phenomenon of inflection. Rich inflectional systems have traditionally been the hallmark of complex grammar. I will give a few examples.

The growth of new function words can be seen nowadays in the emergence of creoles from pidgins. As pidgins and creoles borrow from a superstrate language, the origins of novel function words can be seen quite transparently. For example, in Tok Pisin, the English-based creole of Papua New Guinea, the form bilong functions roughly like the English prepositions of and for, linking nominal expressions as in nambawan pikinini bilong misis kwin ‘first child of Mrs Queen’. The English source word belong is a content word, a main verb, not a function word. Tok Pisin, being a relatively new language, in fact only has two function words, in some analyses, the other one being long, also grammaticalized from an English content word, an adjective as it happens. With fewer function words than, for example, English, it is fair to say that in this respect Tok Pisin has less grammar. It is not so far along the cultural evolutionary scale that leads to languages with greater numbers of
function words. As function words are closely associated with their own peculiar constructions, a language with fewer function words also has fewer distinct grammatical constructions. It seems an inevitable conclusion that the trajectory of complexification of languages has followed a similar path from the earliest beginnings of languages.

The other grammaticalization process that I have identified, namely the rise of inflections from independent words, can also be conveniently illustrated from Tok Pisin. Adjectives, numerals and demonstratives are marked in Tok Pisin by a suffix pela, as in bigpela ‘big’, tupela ‘two’ and dispela ‘this’. This suffix derives from the independent English word fellow. Among non-creole languages, in Romance languages a suffix marking an adverb derived from an adjective, as in French heureusement, is historically derived from the Latin free form mente ‘mind’.

Over the past two millennia, the complexity of inflectional morphology has significantly declined in many languages. Well-known examples are modern English, which has lost most of the inflections of Old English, and the Romance languages, which have lost most of the inflectional complexity of Latin. This might seem to be a case of ‘anti-grammaticalization’, but the mechanisms of growth and loss of inflections are different. The rise of inflection is mediated by phonological erosion of frequently used forms, with the eroded forms being learned as new norms by children acquiring a language. A priority for young children is to conform to the communicative practices of their community, and children quickly learn to mimic adult inflected forms. A different learning situation exists when adults are thrown into strange language communities. Here the first priority is to communicate messages without necessarily conforming to the complex local rules, which are usually too difficult for an adult to master. A Turkish immigrant to Germany needing to buy milk, and knowing the basic content words, can easily get her message across with the incorrect Wo Milch kaufen?, omitting the function words and inflections which would be present in an equivalent request from a native speaker.

To historical linguists of the nineteenth century, the loss of inflections in classical languages was a matter of decline, or loss of perfection. As such, it was a puzzle, unless one were of the pessimistic persuasion that all human behaviour was losing an earlier perfection. There was something of a Garden of Eden myth surrounding the complexities of the classical languages. The puzzle dissolves once one takes a broad view of the course of human history over the past ten millennia. During that recent era, the globe has become much more crowded with humans, who now have to rub up against each other in ways unprecedented during the previous age of isolated hunter-gatherer communities. Long-distance trade and warfare, and the rise of empires and cities inevitably bring adults with different native languages together, and the result is practical forms of communication adopted by adults whose priority is more to get messages across than to conform to the same niceties as their interlocutors. Language contact induces simplification of grammar. Pidgins are the classic and obvious example. Of course, the growth of complexity is not totally stifled, as children still grow up with parents speaking the same somewhat standardized language. But with increasing stirring of the mix of languages in a population, the languages themselves tend to lose the complexity that had been built up over millennia in which society was more stable and less prone to contact with foreigners.

A contrary effect of civilization on language complexity has occurred with the advent of writing, at least in literate people. Writing allows a reader more time to ponder the structure and meaning of a sentence. Hence the quantitative properties of language in use can be tolerably greater in writing than in speech. Sentences and phrases can be longer and there can be more degrees of embedding of subordinate clauses. Educated people start to use
forms in speech that originally would have been characteristic of written language. It is socially impressive to talk like a book.

As in morphosyntax, a cline of complexity can also be seen in phonology. Where pidgins have few or no function words or inflections, they tend likewise to have simple CV syllable structure, and small phoneme inventories avoiding phonetic segments that are typically acquired later in normal acquisition. As with syntax and morphology, complexity of phonology grows over many generations in a stable population not influenced by the simplifying efforts of foreigners.

Summarizing this section, the complexity found in the grammar and phonology of extant languages has evolved by repeated cultural transmission of linguistic norms over many generations, with some reinterpretation by acquirers, leading to a ratcheting effect, so long as the language community is relatively stable. Bringing this whole chapter together, these recent evolutionary developments in complexity of grammar and phonology are built on a much more ancient foundation of cognitive and social evolution, allowing grasp of complex concepts (evolutionary semantics) and negotiation of cooperative social arrangements (evolutionary pragmatics).

Further reading

Over the last twenty-five years, increasingly nuanced views of the origins and evolution of language have developed. This is reflected both in the successive works of individual authors, and in differences between later authors and earlier ones. There is typically now more recognition of the complex interaction and co-evolution of language-related genes and the cultural emergence of complex languages. Inevitably, due to the dominance of the generative paradigm in syntax in the late twentieth century, much writing on language evolution has pivoted around features of Chomsky’s ideas, such as the centrality of syntactic competence, attributed to a genetically determined Universal Grammar (UG), not motivated by function and natural selection. Some authors have consistently argued against such ideas, e.g. Philip Lieberman, who stresses function and phonetic performance. Others have moved over the years from an essentially pro-generative idea that the human syntactic faculty results from catastrophic accidental changes in human DNA to a claim that syntactic ability, and advanced cognition in general, are the inevitable expected outcome of long-term evolution on any planet; this is Derek Bickerton’s intellectual trajectory over more than thirty years, always pursued with great rhetorical vigour. Other authors, especially those writing in the present century, have tended to steer a careful course between the influences of nature and nurture; here I include, in their separate ways, Tecumseh Fitch, Sverker Johansson and myself, James R. Hurford. It is fair to say that writers from within psychology, broadly interpreted, have tended to be strongly opposed to nativist non-functionalist ideas emanating from the dominant paradigm in linguistics; among such psychologists, I include Michael Tomasello, Morten Christiansen and Michael Arbib.

A very useful regular and up-to-date service on recent developments in language evolution is provided by Martin Edwardes at his website named Evolutionary Anthropology Online Research Cluster (EAORC), at http://martinedwardes.webplus.net/eaorc.html. Some of the most salient recent books on language evolution are listed below.

Bibliography

