The Biological Background of Syntax Evolution

Anna Fedor, Péter Ittzés, and Eörs Szathmáry

Abstract

It is difficult to gain an understanding of language since we do not know how it is processed in the brain. Many areas of the human brain are involved in language-related activities, including syntactic operations. Aspects of the language faculty have significant heritability. There seems to have been positive selection for enhanced linguistic ability in our evolutionary past, even if most implied genes are unlikely to affect only the language faculty. Complex theory of mind, teaching, understanding of cause and effect, tool making, imitation, complex cooperation, accurate motor control, shared intentionality, and language form together a synergistic adaptive suite in the human race. Some crucial intermediate phenotypes, such as analogical inference, could have played an important role in several of these capacities. Pleiotropic effects may have accelerated, rather than retarded, evolution. In particular, it is plausible that genes changed during evolution so as to render the human brain more proficient in linguistic processing.

Introduction

Natural language is a fascinating phenomenon, and it is undoubtedly partly biological. Apes, dolphins, and parrots as well as humans are unable to acquire language by learning, no matter how hard they try. There must be something in our genetic endowment that makes humans “ready” for language. Some people think that this readiness is simply due to higher intelligence. Although this may be true in a very broad sense, such claims do not explain how this intelligence differs from that of, say, apes. Humans seem to have gained an insight into the cause and effect in the physical domain, and we are able to produce and use tools by the so-called subassembly strategy. To us it seems that humans possess a few neural procedural capacities that are shared by several important faculties, and that these exist only in very rudimentary forms in other animals. One such procedural element is the ability to handle hierarchical structures efficiently: in the language domain, this is the recursive element of syntax; in
the tool-making domain, this is the subassembly strategy; in the theory of mind
domain, this is second-, third-, and fourth-order intentionality. We propose that
over the past ca. 5 million years, there has been selection on several different
capacities in the hominine lineage; these capacities are partly overlapping, and
the intensity of selection most likely shifted (perhaps several times) over these
domains. There may have been a period during which tool use and tool making
were primarily favored, but then the positively selected genetic variants could
well have turned out to be favorable in some of the other critical domains.

Fisher and Marcus were right when they stated:

In short, language is a rich computational system that simultaneously coordi-
nates syntactic, semantic, phonological and pragmatic representations with each
other, motor and sensory systems, and both the speaker’s and listener’s knowl-
der of the world. As such, tracing the genetic origins of language will require
an understanding of a great number of sensory, motor and cognitive systems, of
how they have changed individually, and of how the interactions between them
have evolved (Fisher and Marcus 2006, p. 10).

The study of language origins has, however, been hampered by the fact that
there is a critical lack of detailed understanding at all levels, including the lin-
guistic one. There is no general agreement among linguists as to how language
should be described: widely different approaches exist, and their proponents
often have very tense scientific and other inter-relationships. As biologists, we
maintain that symbolic reference combined with complicated syntax (includ-
ing the capacity of recursion) constitutes the least common denominator in
this debate. Within this broad characterization, we would like to draw atten-
tion to two approaches which have, perhaps surprisingly, a strongly chemical
character. The first is the Minimalist Program of Noah Chomsky (1995), where
the crucial operator is Merge, the action of which triggers certain rearrange-
ments of the representation of a sentence. There is a broad similarity between
this proposal and chemical reactions (Maynard Smith and Szathmáry 1999).
An even closer analogy between chemistry and linguistics can be detected in
the second approach: Luc Steel’s Fluid Construction Grammar (Steels 2004;
Steels and de Beule 2006). Here, semantic and syntactical “valences” have to
be filled for correct sentence construction and parsing. We should note that
the roots of genetic inheritance are, of course, in chemistry, and that even at
the phenomenological level Mendelian genetics was a stoichiometric para-
digm, influenced by contemporary chemical understanding (elementary units
that can be combined in certain fixed proportions give rise to new qualities).
Chemical reactions can be also characterized by rewrite rules. In-depth study
is required to consider the ramifications of this analogy: the deeper it goes, the
more benefit one can hope from taking the analogy seriously.

There are now attempts to rethink Fluid Construction Grammar in terms of
replicator dynamics within the brain. This may sound surprising, but one should
not forget that the question of whether thinking or language is performed in the
brain by processes analogous to evolution by natural selection is wide open. Ever since the ideas of William James (1890), some suspect that brain dynamics will be proven to include a crucial evolutionary element, just as it is now known to apply for the immune system. The immune system is particularly interesting because it is both an evolutionary (D’Eustachio et al. 1977) and generative system (Jerne 1985). A few selectionist approaches to brain epigenesis and function exist (Dawkins 1971; Changeux 1983; Finkel and Edelman 1985) but they lack one crucial component: multiplication. The only existing attempt was carried out by Calvin (1996), and this triggered further research on how language may be processed in the brain (Calvin and Bickerton 2000). The main problem with Calvin’s mechanism is the lack of connectivity copying from one brain site to another. Such a mechanism is crucial if we think that connectivity encodes information about alternative hypotheses, among which some reward mechanism selects the better ones (Fernando et al. 2008), and it is also needed to close the gap between neuroscience and the evolutionary epistemology of, say, Campbell (1974).

Language needs certain prerequisites. Some, however, are not especially relevant to the main problems addressed in this volume. For example, apes do not possess a descended larynx nor do they have cortical control over their vocalizations. Undoubtedly, these traits must have evolved in the human lineage, but we do not think that they are indispensable for language as such. One could have a functional language with a smaller number of phonemes, and sign language does not require either vocalization or auditory analysis (Senghas et al. 2004; MacSweeney et al. 2008). Thus, our focus is primarily on the neuronal implementation of linguistic operations, irrespective of the modality. It is difficult to imagine the origin of language without capacities for teaching (which differs from learning), imitation, and a complex theory of mind (Premack 2004). Apes are limited in all these capacities. It is fair to assume that these traits have undergone significant evolution because they were evolving together with language in the hominine lineage. To this one should add—not as a prerequisite, but as a significant human adaptation—the ability to cooperate in large non-kin groups (Maynard Smith and Szathmáry 1995). Together, these traits form an adaptive suite that is specific to humans. We suggest that in any selective scenario, capacities for teaching, imitation, some theory of mind, and complex cooperation must have been rewarded, because an innate capacity for these traits renders language evolution more likely (Szathmáry and Számadó 2008b).

From the neurobiological perspective, we call attention to the fact that some textbooks (e.g., Kandel et al. 2000) still present a distorted image of the neurobiological basis of language. It is very simplistic to assign the Wernicke and Broca areas of the left hemisphere to semantics and syntax, respectively. The localization of language components in the brain is extremely plastic, both between and within individuals (Neville and Bavelier 1998; Müller et al. 1999). Surprisingly, if removal of the left hemisphere happens in the first few months
after birth, the patient can nearly completely retain his/her capacity to acquire language. This stands, of course, in sharp contrast to the idea of anatomical modularity. It also severely limits the idea that only the afferent channels changed during the evolution of the human brain: modality independence and the enormous brain plasticity in the localization of language favor the idea that whatever has changed in the brain to render it capable of linguistic processing must be a very widespread property of the neuronal networks (Szathmáry 2001). Components of language (as well as those of other capacities) get localized during development somewhere in any particular brain in the most functionally “convenient” parts available (cf. Karmiloff-Smith 2006). Language is just a certain activity pattern of the brain that finds its habitat, like an amoeba in a medium. The metaphor “language amoeba” expresses the plasticity of language, but it also calls attention to the fact that a large part of the human brain is apparently a potential habitat for it; no such habitat appears to exist in nonhuman ape brains (Szathmáry 2001).

For a long time, there has been a dogma concerning the histological uniformity of homologous brain areas in different primate species. Recent investigations, however, do not support this claim (DeFelipe et al. 2002). In fact, the primary visual cortex shows marked cytoarchitectonic variation (Preuss 2000), even between chimps and man. Therefore, one cannot at all exclude the possibility that some of the species-specific differences in brain networks are genetically determined, and that some of these are crucial for human language capacity. As discussed above, these language-critical features must be a rather widespread network property. Genes affect language through the development of the brain. Thus the origin of language must be to a large extent an exercise in the linguistically relevant developmental genetics of the human brain (Szathmáry 2001).

Consider the existing data on genetic changes that are more directly relevant to language. The FOXP2 gene was discovered to have mutated in an English-speaking family (Gopnik 1990, 1999). It has a pleiotropic effect: it causes orofacial dyspraxia, but it also affects the morphology of language. Affected patients must learn or form the past tense of verbs or the plurals of nouns case by case, and even after practice they do so differently from unaffected humans (for a review, see Marcus and Fisher 2003). FOXP2 underwent positive selection (Enard et al. 2002) in the past, which demonstrates that there are genetically influenced important traits of language other than recursion (Pinker and Jackendoff 2005), contrary to other opinions (e.g., Hauser et al. 2002). In addition, there is a known human language that apparently has no recursion: the Pirahã language in the Amazon (Everett 2005). It would be good to know how these particular people manage recursion in other domains, such as object manipulation or “action grammar” (cf. Greenfield 1991).

The capacity to handle recursion appears to differ from species to species. Tamarin monkeys, for example, have demonstrated insensitivity to auditory patterns defined by more general phrase structure grammar, but they discover
violations of input conforming to finite state grammar (FSG; Fitch and Hauser 2004). It would be interesting to know how chimps would perform in this same experiment. Human adults are sensitive to both violations (cf. below). Needless to say, it would be very interesting to know the relevant sensitivities in apes and human children (preferably just before they master grammar of natural language). One should design an experiment capable of producing consistent patterns of such a capacity in evolving neuronal networks, and then reverse engineer proficient networks to discover the evolved mechanisms for this capacity (see below).

We share the view that language is a complex, genetically influenced system for communication that has been under positive selection in the human lineage (Pinker and Jackendoff 2005). The task of the modeler, then, is to try to model intermediate stages of a hypothetical scenario and, ultimately, to re-enact critical steps of the transition from protolanguage (Bickerton 1990) to language. It cannot be denied that language is also a means for representation. This is probably most obvious for abstract concepts, for which the generative properties of language may lead to the emergence of a clear concept itself. This is well demonstrated for arithmetics. For instance, an Amazonian indigenous group lacks words for numbers greater than 5; hence they are unable to perform exact calculations in the range of larger numbers, but they do have approximate arithmetics (Pica et al. 2004).

Language changes while the genetic background also changes (this must have been true especially for the initial phases of language evolution), and the processes and timescales involved are interwoven. This opens up the possibility for genetic assimilation: some changes that each individual must learn at first can later become hardwired in the brain. Some have endorsed the importance of this mechanism in language evolution (Pinker and Bloom 1990), whereas others have raised doubt (Deacon 1997). Deacon’s argument against it is that linguistic structures change so fast that there is no chance for the genetic system to assimilate any grammatical rule. This is true, but not very important. There are linguistic operations—performed by neuronal computations and related, among others, to compositionality and recursion—that must have appeared sometime in evolution. Whatever the explicit grammatical rules are, such operations must be executed.

Hence, a much more likely scenario for the importance of genetic assimilation proposes that many operations must have first been learned, and those individuals whose brain was genetically preconditioned to a better (faster, more accurate) performance of these operations had a selective advantage (Szathmáry 2001). Learning was important in rendering the fitness landscape more climbable (Hinton and Nowlan 1987). This view is consonant with Rapoport’s (1990) view of brain evolution. This thesis is also open for experimental test.

The origin of language is an unsolved problem; some refer to it as the “hardest problem of science” (Christiansen and Kirby 2003). What makes it difficult is the fact that physiological and genetic experimentation on humans and
apes is very limited. The uniqueness of language prohibits, strictly speaking, application of the comparative method, so infinitely useful in other branches of biology. Fortunately, some components of language lend themselves to a comparative approach, as we shall see in relation to birdsong. Nevertheless, limitation of the approaches calls for other types of investigation.

We believe that simulations of various kinds are indispensable elements of a successful research program. Yet a vast range of computational approaches have brought less than spectacular success (cf. Elman et al. 1996). In our opinion, this is attributable to the utterly artificial nature of many of the systems involved, such as connectionist networks using back-propagation, for example (for a detailed criticism, see Marcus 1998).

In this chapter, we review some key findings and ideas concerning the genetic, neurobiological, and evolutionary background of the “language problem.” In addition, we provide an update on some of some previous suggestions, present our model for a minimalist neural network parsing phrase/structure grammar, and discuss arguments in favor of a human-specific adaptive suite.

### Genetic Background of Language

Information about the human and the chimp genome (Chimpanzee Sequencing and Analysis Consortium 2005) is now “complete,” and one can ask how far previous optimism seems justified in light of comparative studies based on this information. It is clear that much work lies ahead. Knowing all the genes of chimps and humans is not everything: we need to know how the genotype is mapped to the phenotype, and this is a formidable problem. Genes are expressed in specific ways, under the influence of other genes and the environment. Interaction between genes is not the exception but the rule. One gene can affect several traits (pleiotropy) while actions of different genes do not affect traits (including fitness) independently (epistasis). It is the network of interactions that is of importance, and one must not forget that there are networks at different levels: from genetic regulatory networks through protein interaction networks and signal transduction pathways to the immune system or neuronal networks. The question is how the effect of genes percolates upwards. Genes act on expressed molecules (proteins and RNA) that do their job in their context. There is something amazing about the fact that hereditary action on such primitive molecules percolates upwards, resulting in heritability of complex cognitive processes, including language.

The chimp and human genomes are indeed similar, but one should understand clearly what this means (Fisher and Marcus 2006). Substitutions make up for 1.23% of difference between the two genomes; this translates into 35 million altered sites in the single-copy regions of the genomes! Insertions and deletions yield a further 3% genomic difference. It is convenient to distinguish between altered structural and regulatory genes. The first codes for altered
enzymes or structural proteins; the latter codes for altered transcription factors, for example. Both kinds of changes happened since humans diverged from chimps, and both affected language in critical ways.

There seems to have been an acceleration in the changes of neural gene expression patterns in human evolution, although this should be evaluated against the background that liver and heart expression patterns have diverged a lot more between chimps and humans. The usual interpretation is that neural tissue is under stronger stabilizing selection. Another observed tendency is the up-regulation of human neural gene expression relative to the chimp, but the functional significance of this finding is unclear (it may be a more or less direct result of recent genomic region duplications).

It is not yet clear what gene expression differences exist behind the cytoarchitectonic differences among the Brodmann areas: the most known differences between chimps and humans are common to all cortical regions. Recently, this view has been refined. Oldham et al. (2006) analyzed gene co-expression patterns in humans and chimps and were able to identify network modules that correspond to gross anatomical structures including the cerebellum, caudate nucleus, anterior cingulated cortex, and cortex. The similarity of network connectivity between the respective human and chimp areas decreased in that order, consistent with the radical evolutionary expansion of the cortex in humans. It is intriguing to note that in the cortical module there is a strong co-expressive link between genes of energy metabolism, cytoskeletal remodeling, and synaptic plasticity.

There are genetic changes that probably did boost language evolution but in a general, aspecific way. Genes influencing brain size are likely to have been important in this sense. Note, however, that genes involved in primary microcephaly seem to have been under positive selection in the past, but children with this syndrome can have rather normal neuroanatomical structures despite the fact that their overall brain size may be reduced to a mere one third of the normal. They show mild to moderate mental retardation but pass several developmental stages. Fisher and Marcus conclude:

In our view the honing of traits such as language probably depended not just on increased “raw materials” in the form of a more ample cortex, but also on more specific modification of particular neural pathways (Fisher and Marcus 2006, p. 13).

Perhaps the most revealing recent finding concerning genetic brain evolution is the identification of an RNA gene that underwent rapid change in the human lineage (Pollard et al. 2006). It is expressed in the Cajal-Retzius cells of the developing cortex from 7 to 19 gestational weeks. It is co-expressed with reelin, a product of the same cells, which is important in specifying the six-layer structure of the human cortex.

Even if some of our linguistic endowment is innate, there may not be much genetic variation for the trait in normal people, just as most people have ten
fingers. In contrast, our linguistic capacity may be like height: whereas all people have height, there are quantitative differences in normal people. To be sure, children as well as adults differ in their linguistic skills. However, to what extent do genes account for this variation?

Surveying many studies, Stromswold (2001) concluded that twin concordance rates are significantly higher for monozygotic than for dizygotic twins. Twins are concordant for a trait if both express the trait or if neither expresses it. Twins are discordant for a trait if one exhibits the trait and the other does not. If the concordance rate for language disorders is significantly greater for monozygotic than dizygotic twins, this suggests that genetic factors play a role in language disorders such as dyslexia and specific language impairment (SLI). The concordance rates for written and spoken language disorders are similar. For both written- and spoken-language disorders, the mean and overall concordance rates were approximately 30% higher for monozygotic than for dizygotic twins, and genetic factors accounted for between one-half and two-thirds of the written and spoken language abilities of language-impaired people. In studies of normal twins, depending on the aspect of language being tested, between one-quarter and one-half of the variance in linguistic performance was attributable to genetic factors. People have been tested on phonological short-term memory, articulation, vocabulary, and morpho-syntactic tasks. It seems that different genes may be responsible for the variance in different components and language and that some genetic effects may be language-specific.

The sum of all genetic effects is usually not much greater than 50% for various aspects of cognition (Stromswold 2001). Most individual genes are expected to have small effects. Candidate genes affect functions such as the cholinergic receptor, episodic memory, dopamine degradation, forebrain development, axonal growth cone guidance, and the serotonin receptor. It is a great problem that cognitive skills are likely to have been, at least in part, inadequately parsed; thus so-called intermediate phenotypes with a clearer genetic background should be sought. By this token, schizophrenia as such does not exist; rather, different genes may go wrong and the symptoms such as hallucinations are emergent outcomes (Goldberg and Weinberger 2004). The situation may be similar to that of geotaxis in Drosophila, where the individual involvement of different genes that collectively determine this capacity is counterintuitive (Toma et al. 2002).

It is worth calling attention to the fact that the genetics of human cognitive skills is a notoriously difficult problem. One common reason is that usually the clinical characterizations are not sufficient as descriptions of phenotypes (Flint 1999). A consensus seems to emerge that the genes involved are so-called “liability genes” which, when present in the right allelic form, significantly enhance the probability of developing the respective cognitive skills.

Perhaps the most important neurodevelopmental syndrome for our topic is SLI, where there is significant difference between verbal and nonverbal skills.
Several candidate chromosomal regions have been identified (SLI consortium 2002).

A by now famous gene is FOXP2, first identified by Gopnik (1990). In a certain English-speaking family, a dominant allele was found to cause the syndrome developmental verbal dyspraxia (DVD), formerly grouped under SLI. No one disputes the fact that SLI is real. What is contested is how closely it is limited to, or rooted in, a specific grammatical impairment. The Gopnik (1990, 1999) case has been very stimulating because of its characterization as “feature-blind” dysphasia and its obvious genetic background (a single dominant allele); however, cognitive skills are affected as well (Vargha-Kadem et al. 1998). More evidence with other linguistic groups is accumulating (Dalalakis 1999; Rose and Royle 1999; Tomblin and Pandich 1999). One study (Van der Lely et al. 1998), sadly without genetics, claims to demonstrate that grammatically limited SLI does exist in “children” (although only one child is analyzed in the paper).

The FOXP2 protein is an old transcription factor present in vertebrates, and there is evidence that it has been under positive selection in the human lineage. It seems to affect development of distributed neural networks across the cortex, striatum, thalamus, and cerebellum. DVD differs from SLI, but speech and language deficits are always present, even in otherwise normal children. In other affected individuals, general intelligence is impaired and grammar deficits (difficulty with morphological features such as the suffix –s for plural or –ed for past tense) occur in written language as well. The selective sweep that affected this gene in the human lineage occurred within the past 200,000 years (Enard et al. 2002; Zhang et al. 2002).

Analysis of the expression patterns of FOXP2 in other species suggests that this gene has been involved in the development of neural circuitry processing sensorimotor integration and coordinated movements, lending support to the notion that language has its roots in motor control (e.g., Lieberman 2007). This makes the involvement of basal ganglia in speech and language less than surprising.

Recent studies (reported by White et al. 2006) demonstrate that FoxP2, although without accelerated evolution, plays a crucial role in the development and seasonal activation of relevant brain areas in songbirds. Interestingly, although the avian and human forms are very similar, neither of the human-specific mutations has been found in the FoxP2. Also of interest is the fact that the ganglia involved in birdsong learning seem to be analogous to the basal ganglia involved in human vocal learning (Scharff and Haesler 2005).

Researchers have called attention to the fact that in songbirds and humans both FoxP2 and FoxP1 are expressed in functionally similar brain regions that are involved in sensorimotor integration and skilled motor control (Teramitsu et al. 2004). Moreover, differential expression of FoxP2 in avian vocal learners is correlated with vocal plasticity (Haesler et al. 2004). Using songbirds as
analogs to human learning of speech, Haesler et al. (2007) proved that birds with FoxP2 knockout suffer from incomplete and inaccurate vocal imitation.

Mice, like humans, have two copies of the Foxp2 gene as well. If only one of them is affected, pups are severely affected in the ultrasonic vocalization upon separation from their mother. This suggests a role for this gene in social communication across different species. The Purkinje cells in the cerebellum are affected in the pups (Shu et al. 2005). Determination of the expression pattern in the developing mouse and human brain is consistent with these investigations: regions include the cortical plate, basal ganglia, thalamus, inferior olives, and cerebellum. Impairments in the sequencing of movement and procedural learning may thus underlie the linguistic symptoms in humans (Lai et al. 2003). According to Vernes et al. (2007), the targets of this regulatory gene in mice include loci involved in modulating synaptic plasticity, neuronal development, axon guidance, and neurotransmission. Spiteri et al. (2007) identified transcriptional targets of FOXP2 in human basal ganglia and the inferior frontal cortex. Many target genes play roles in neurite outgrowth and plasticity. Fujita et al. (2008) inserted a human-specific FOXP2 gene into mice, which in homozygous condition die early, have (among other impairments) abnormal Purkinje cells, and show severe ultrasonic vocalization and motor impairment.

Recently, Krause et al. (2007) attempted to date the fixation of the two human-specific amino acid substitutions in FOXP2 by claiming that the gene was shared by Neanderthals and hence the substitutions had been fixed more than 300 thousand years ago. This conclusion was challenged by Coop et al. (2008) on methodological grounds, so the jury is still out on this issue.

It is important to emphasize that the link between genes and mental capacities is extremely indirect (e.g., Karmiloff-Smith 2006): genes encode for RNA and protein molecules, and every effect on behavior must penetrate “upwards” through a large molecular and cellular interaction network. Williams syndrome provides a good case in point. These patients are regarded as handicapped in spatial orientation and yet are good at language: 28 contiguous genes are involved in the phenomenon. However, a closer analysis of their language reveals that it is not “normal” either, and it also develops late in life. Even a mouse “model” exists with a mutant LIMK1 (a protein kinase gene expressed in the developing brain) for the spatial problem. This gene is expressed not only in regions responsible for spatial orientation, and some human patients have an impaired LIMK1 gene yet they do not exhibit Williams syndrome. By the same token, many children with SLI have no problem in the FOXP2 gene. Here it is instructive to quote from Karmiloff-Smith (2006, p.15):

WS [Williams syndrome] is caused by a deletion of some 28 genes on one copy of chromosome 7; D(own) S(yndrome) is caused by an additional whole chromosome 21; Fragile X is caused by a mutation of a single gene on the X chromosome; velocardiofacial syndrome (or di George syndrome) is caused by a large deletion on chromosome 22. Yet all four syndromes display both delay and
deviance, mental retardation, gross and/or fine motor deficits, impaired sleep patterns, memory deficits, number impairments, and often hyperactivity. Three of them show better language skills than spatial skills.

We believe that the biologically motivated dissection of the language faculty is of primary importance. Put differently: What are the intermediate phenotypes that make up language? This question cannot be answered, we believe, without an appropriate formulation of aspects of language. Thus linguistic theories must ultimately be biologically constrained. A good start in this direction may be Fluid Construction Grammar (Steels and de Beule 2006). To date, though, there is not much coupling of details of linguistic theories to those of brain mechanisms.

Brain and Language

The analysis of neural activity during the performance of cognitive tasks has become a burgeoning industry. Sensitivity has increased over the years, and these methods have been increasingly applied to the recording of brain activity during linguistic performance. The crucial observation to bear in mind is that localization of certain functions to particular brain areas during normal development does not necessarily mean that the particular region is a “hard-wired region in the modular sense” for that particular function. Brain development, especially in the first few months in life, is very plastic and many cognitive skills can be (nearly) spared due to plastic recovery after early injury. The same applies to components of language, even syntax. One can surely learn about the neurobiological foundations of syntax by studying Broca’s area in normal people or patients with late lesion of that area, but at the same time one should also ask how the relevant tasks are performed in patients who do not have Broca’s area at all!

Where Is Language in the Brain?

The recognition that neural localization of language can be plastic is widely known (Nobre and Plunkett 1997; Neville and Bavelier 1998; Musso et al. 1999). Studies of brain injury have shown that when damage to the left hemisphere is sustained before a critical period, the right hemisphere is able to take over the necessary functions (Müller et al. 1999). This does not contradict the finding that in normal people Broca’s area seems specialized for syntax (Embick et al. 2000). It appears that the common left-hemisphere localization of language is just the most likely outcome in the absence of genetic or epigenetic disturbance. What is more, both the cortical and subcortical areas contribute to language processing; reward systems and motor control provided by basal ganglia and the cerebellum seem to be critical components of our language faculty (Lieberman 2002, 2007).
The conclusions that we can draw from brain studies are as follows:

- Localization of language is not fully genetically determined: even large injuries can be tolerated before a critical period.
- Language localization to certain brain areas is a highly plastic process, both in its development and end result.
- A surprisingly large part of the brain may sustain language: there are (traditionally recognized) areas that seem to be most commonly associated with language, but they are by no means exclusive, either at the individual or the population level, during either normal or impaired ontogenesis.
- Whereas a large part of the human brain can sustain language, no such region exists in apes.
- Language processing has a distributed character.

It is instructive to look at the evolutionary patterns of the sensory neocortex in mammals (Krubitzer and Kaas 2005). Auditory, somatosensory, and visual fields (contiguous brain tissue regions) have changed in location and size in different species. Fields can change in absolute and relative size, as well as in number. Connections of cortical fields can also change. Such alterations can be elicited by manipulation of either the peripheral morphology or activity, or that of the expression level of certain genes. Phenotypic within-species variation can be extremely broad; however, little is known about the relative magnitude of the genetic part of this variation. A good example of genetic influence is the variation in the cortical area map of inbred mice, reflecting strain identity (Airey et al. 2005).

Evolution of the vertebrate brain has produced an increase in cortical size and elaboration of the cortical circuit diagram (Hill and Walsh 2005). Most importantly, cortical layers II and IIIb, IIIc of the chimp differ from layers IIa, IIb and IIIa, IIIb and IIIc, respectively, in humans. A tentative conclusion, based on “rewired” ferrets and three-eyed frogs, is that layers form independently of patterned input, and instructive electrical signals play a crucial role in fine network development, which also affects intracortical connections (Sur and Learney 2001).

Genetically determined patterning of parts of the brain follows mechanisms well-known from conventional developmental studies. For example, during the formation of the retinotopic map, axons from the retinal ganglion cells find their targets in the tectum as a result of matching between two receptor/ligand pairs (Schmitt et al. 2006), both expressed according to (altogether four) gradients, (two in the eye and two in the tectum).

Several people, including Greenfield (1991), have suggested the involvement of tool making in the evolution of language, for example, in the form of “action grammar” which can be recursive when agents use the “subassembly” strategy in the “nesting cups” experiment. The idea is that selection for efficient tool use could have aided language evolution and vice versa. Stout and
Chaminade investigated via brain imaging the neurobiological bases of this in modern naïve (i.e., untrained) humans by requiring them to implement a 2.5 million year-old Oldowan practice of tool making. (Incidentally, in that material culture we see evidence for the uniquely human practice of using a tool to make another tool.) Premotor cortex was activated in the task but not the prefrontal executive cortex (involved in planning) nor the inferior parietal cortex. The activation of caudal Broca’s area in this task underlines the possible link between language and tool making, and is consistent with views of the importance of “mirror neurons” in language evolution (Rizzolatti and Arbib 1998)—the latter, however, are by no means sufficient for language, as many animals possess them. As we learn words by imitation, and tool-making requires an “action grammar,” it is unlikely to be accidental that human Broca’s area evolved from structures that are involved in these capacities beyond (and prior to) language. Caplan (2006b) suggests that Broca’s area is involved in syntactic processing, not merely because it is evolutionarily related to the dorsolateral prefrontal cortex or its original involvement in sensorimotor functions, but because of its intrinsic neural organization. However, this leaves the very essence of the suggested neural organization obscure.

Grodzinsky and Santi (2008) present evidence in favor of the view that Broca’s area is specifically involved in syntactic movement (for a discussion of what Movement means, see Bickerton and Rizzi, both this volume) rather than syntactic complexity per se, although they also accept that it is involved in language production and working memory (Friederici, this volume). Note that this conclusion is drawn from analysis of either normal people or patients in which the right hemisphere has not taken over language processing. Thus, once again, Broca’s area is exciting but by no means exclusively so. The lesson, however, is to figure out in which way could Movement require specific neuronal mechanisms relative to other linguistic operations.

A Minimalist Neural Network Parsing Phrase-Structure Grammar

A crucial element of syntax is center-embedded recursion (Hauser et al. 2002), which has been regarded as specific to humans. This view was recently challenged by Gentner et al. (2006), who believe to have demonstrated that European starlings recognize context-free grammar (CFG; Figure 2.1) with center-embedding. This experimental design was influenced by the former experiment of Fitch and Hauser (2004), who interpreted their results as showing that, whereas tamarin monkeys as well as human students recognize FSG, only humans recognize CFG.

The methodological problem with these studies is that because there is no need for real center-embedding (bracketing), the task can be solved by counting (Corballis 2007a, b). As Corballis aptly wrote:
To distinguish between these models, information beyond the strings themselves is required. That information might be semantic, or prosodic, or perhaps even some neurophysiological process that makes embedding mandatory. It might also depend on associative learning. For example, \( A^n B^n \) strings might be parsed as center-embedded if there were established associations between \( AB \) pairs from the outside in—such that there is an association between the first and last, second and second to last, and so on (Corballis 2007b, p. 1582).

Consonant with this approach is the experimental finding that humans also perform poorly on learning center-embedded structures when other methods (such as counting) are not allowed in artificial grammar learning (de Vries et al. 2008).

Sun et al. (1998) implemented a hybrid system in which a recurrent neural network was coupled to an external nonneural stack memory. After training with backpropagation, the system was able to infer a CFG from input. In another study (Bodén and Wiles 2000), continuous time recurrent networks without a stack can learn both context-free (\( A^n B^n \)) and context-sensitive (\( A^n B^n C^n \)) languages in a prediction task, using backpropagation through time. Since there were no long-range dependencies connecting words within the sentences, performance of these systems boiled down to counting (Rodriguez et al. 1999). Chen and Honavar (1999) proposed an artificial neural network architecture for syntax analysis through the systematic composition of a suitable set of component symbolic functions realized using neural associative processor modules. The neural associative processor is a 2-layer perceptron that can store and retrieve arbitrary binary pattern associations. Their model is a fairly complex system which can avoid the problem mentioned above.

Recently, we performed a study to examine these issues further (Fedor et al., submitted). Our aim was to handcraft a minimalist neural model that can parse real center-embedded structures with established associations between \( AB \) word pairs as mentioned above (Figure 2.1, right panel). Although the proposed network is not directly biologically realistic, we believe it can be smoothly transformed into such an architecture. We rely on the observation

![Figure 2.1](image-url)  
**Figure 2.1** Apparent (left) and real (right) context-free grammar. The left structure can be parsed by simple counting; the tree on the right needs some knowledge of the context-free grammar because of the long-range dependencies (word pairs). After Corballis (2007a, b).
that CFG requires *some* implementation of a stack, with the necessary *pop* and *push* operations (Hopcroft and Ullmann 1979). The task is then to come up with a near-realistic and minimalist neural architecture.

Our proposed network is simpler than the above solutions and avoids backpropagation. It rests on the assumption that gating of synaptic connections is critical for complex cognitive processes (e.g., Gisiger et al. 2005; O’Reilly 2006). There are four main components of the neural network: the input layer, the clocked stack, the pairing module, and the end-of-sentence neuron (Figure 2.2). The input layer receives one word at a time from the sentence. The stack consists of a number of layers, each capable of storing one word at a time. The task of the pairing module is to compare two words: one on the input layer and the other on the top of the stack. The result is given to the push-pop neuron (PP), which signals 1, if the words are pairs, and signals –1, if the words are not pairs. The PP neuron performs gating on the inter-layer synapses of the stack. It is connected to every upward synapse with weights of 1 and to every downward synapse with weights of –1. As a result, if the words are pairs, it inhibits the downward connections in the stack; hence upward connections will predominate, and each layer will take the value of the layer below it (a pop action). In this case, the bottom layer will take the value of the empty sign. If, however, the words are not pairs, the PP neuron inhibits the upward

---

**Figure 2.2** Architecture of the grammar parsing network. The stack is depicted by the gray background. Neurons (N) are part of the pairing module that compares the input with the top of the stack. The PP (push/pop) neuron blocks or activates the different synapses of the stack and in this way either pushes down the next input or pops up the word from the top of the stack. The end-of-sentence neuron is not depicted in the figure. Only one word from the input sentence is represented here.
connections in the stack, and downward synapses will predominate such that every layer will take the value of the layer above it (a push action). In this case, the top layer takes its next value from the input. Finally, there is an end-of-sentence neuron, which signals 1 only if there is the end-of-sentence sign on the input. The stack should be empty at the same time, if the sentence was grammatically correct; if the (random) sentences are grammatically incorrect, some words get stuck in the stack.

This architecture was tested with input conforming to FSG and CFG. Each synaptic weight was trained independently with the simple perceptron learning rule during several learning sessions until weights converged. After learning, testing was performed with a mix of novel CFG, FSG, and randomly generated sentences, and the performance of the network was scored. Two types of performance were measured: the percentage of word pairs recognized and the percentage of sentences correctly categorized to grammatical and nongrammatical. The recognition of word pairs is irrespective of stack depth, and the network can learn word pairs correctly provided that every possible word pair was present in the learning set. Sentence categorization, however, is dependent on the depth of the stack: to parse an \( n \)-word-long CFG sentence successfully, an \( n \)-layer-deep stack is required, whereas to parse an FSG sentence, a 1-layer-deep stack is always enough.

Apart from the FSG and CFG grammars mentioned above, the network was exposed to another type of CFG, called the palindrome. These types of sentences are similar to the center-embedded structure, but bounded words are identical, not pairs. The network can learn palindrome grammar with equal ease, provided that the depth of the stack is at least half the number of the words in the palindrome sentence. Of course, if a network is trained on palindrome sentences, it will not recognize the original FSG and CFG grammars as correct, and vice versa.

The stack implemented here follows a design embedded in the chemical literature (Hjelmfelt et al. 1992) that rests on gating. We strongly believe that gating will be found to be crucial for hierarchical tasks, just as for complex cognition in general (Gisiger et al. 2005; O’Reilly 2006). The fact that it has readily evolved in a reinforcement-learning task in a simulated honeybee neural network (Soltoggio et al. 2007) supports this idea. The design of our minimalist network is very pragmatic in that it includes a perceptron (the pairing module) which can be substituted by a more realistic neuronal network if necessary. In contrast, we suggest that the introduction of the neural stack memory (pushdown automaton) will turn out to be substantial for any biological “hierarchical processor.” The performance of our network naturally depends on the depth of the stack, and as such it can be replaced by a finite state automaton (Hopcroft and Ullman 1979). However, in this sense, human parsing ability is also limited: no person can parse sentences with arbitrarily many hierarchical layers (Pinker 1994). The likely hierarchical processor (maybe even supramodal) in humans with normal development is Broca’s area (Friederici 2006; Tettamanti
and Weniger 2006). Sadly, we know next to nothing about the relevant “internal wiring” of this area: we propose that it is likely to contain a neuronal stack, wherein gating will be found important.

**Brain Epigenesis and Gene–Language Coevolution**

It has to be admitted that on the whole we do not understand how the brain works. Nevertheless, some crucial elements seem to emerge. One is that development of the normal brain is enormously plastic, even though the power of genetic factors is obvious. One classic example is that in the same brain areas of identical twins, the two hemispheres of the same individual resemble each other more closely than the same hemispheres in the two people (Changeux 1983).

Another insight is that a tremendous amount of variation and selection transpires during brain ontogenesis. This is a Darwinian-type process, no doubt. As William James recognized a long time ago, natural selection of heritable variation is the only known force that can lead to adaptations, so let’s apply it to brain ontogenesis and problem-solving as well (James thought that even learning is the result of selection of variation within the brain). There are several expositions that all regard the brain, one way or the other, as a “Darwin machine” (Calvin 1987). Here, for simplicity, we stick to the formulation by Changeux (1983), who stated that the functional microanatomy of the adult cortex is the result of the vast surplus in initial stock of synapses and their selective elimination according to functional criteria (performance).

In the previous section we learned that a very large part of the human brain can process linguistic information, including syntactical operations. This means that there is no fixed macro-anatomical structure that is exclusively dedicated to language, but some functional micro-anatomical structure *must* be appropriate, otherwise it could not sustain language. This further suggests that there is some *statistical connectivity feature* of a large part of the human brain that renders is suitable for linguistic processing (Szathmáry 2001). From the selectionist perspective there are three options: the initial variation in synaptic connectivity is novel; the means of selection on functional criteria is novel; or both. Maybe both component processes are different in the relevant human brain areas, and we do not dare to speculate about their relative importance.

This idea must be seen in close connection to the one presented by Rapoport (1990) concerning the coevolution of brain and cognition (within a human a population of humans). The traditional view is the so-called bottom-up mechanism: that is, a genetic change of some neural structure is subjected to selection and, based on its performance, it either spreads or it does not. There is, however, a so-called top-down mechanism, which could have contributed more significantly to the evolution of human cognitive skills, including, especially we argue, language. The crucial idea is as follows:
Due to the plasticity in brain development, enhanced demands on a certain brain region lead to less synaptic pruning (a known mechanism). Less synaptic pruning is assumed to lead to more elaborate (and more adaptive) performance. Any genetic change contributing to the growth of the brain area thus affected will be favored by natural selection.

Two important connections deserve to be highlighted. First (observed by Rapoport himself), the top-down mechanism is a more detailed exposition of the late Allan Wilson’s idea (Wyles et al. 1983). Thus an increased brain, due to its more complex performance, alters the selective environment (in social animals composed of conspecifics to a great extent), which selects for an even larger brain, and so on. Second, and perhaps more important, this mechanism is also a neat example of a Baldwin effect (or genetic assimilation), when “learning guides evolution.” As Deacon (1997) pointed out, it is trickier to apply the idea of genetic assimilation to language than usually thought. The reason for this is that the performed behavior must be sufficiently long lasting and uniform in the population. It is thus hard to imagine how specific grammatical rules, for example, could have been genetically assimilated. This point is well taken, but here we speak of a different thing: the genetic assimilation of a general processing mechanism that is performed by virtue of the connectivity of the underlying neural structures.

Our claim is that the most important, and largely novel, faculty selected for was the ability of the networks to process syntactical operations on symbols that are part of a semantically interwoven network. The specific hypothesis is that linguistically competent areas of the human brain have a statistical connectivity pattern that renders them especially suitable for syntactical operations. In conclusion, we think that:

- The origin of human language required genetic changes in the mechanism of the epigenesis in large parts of the brain.
- This change affected statistical connectivity patterns and dynamical development of the neural networks involved.
- Due to the selectionist plasticity of brain epigenesis, coevolution of language and the brain resulted in the genetic assimilation of syntactical processing ability as such.

An intriguing possible example of gene–culture coevolution has recently been raised by Bufill and Carbonell (2004), who call attention to a number of facts. First, human brain size has not increased over the past 150,000 years; in fact, it has decreased somewhat in the last 35,000 years. Second, a new allele of the gene for apolipoprotein E (ApoE4) originated sometime between 220,000 and 150,000 years ago. This allele improves synaptic repair (Teter et al. 2002). The original form entails a greater risk of Alzheimer disease and a more rapid, age-related decline in general (Raber et al. 2000). More importantly, ApoE4 impairs
hippocampal plasticity and interferes with environmental stimulation of synaptogenesis and memory in transgenic mice (Levi et al. 2003). Interestingly, the ancestral allele decreases fertility in men (Gerdes et al. 1996). Taken together, these facts indicate, but do not prove, a role in enhanced synaptogenesis during a period when syntactically complex language is thought to have originated. More evidence like this would be welcome in the future, since one such case can at best be suggestive.

Selective Scenarios for the Origin of Language

On the Human-specific Adaptive Suite

Various people (e.g., Premack 2004) have called attention to that fact that besides language, efficient teaching (which differs from learning), imitation, and a developed theory of mind are also uniquely human resources. We would add to this the trait of human cooperation (Maynard Smith and Szathmáry 1995), which is remarkable in that humans are able to cooperate even in large non-kin groups. We propose that these traits did not appear by accident together. They form an adaptive suite, and presumably they have coevolved in the last five million years in a synergistic fashion (Szathmáry 2008; Szathmáry and Számadó 2008b). A relevant image is a coevolutionary wheel (Figure 2.3): evolution along any of the radial spokes presumably benefited all the other capacities, even if the focus of selection may have changed spokes several times. This hypothesis is testable; and there is indeed already evidence in its favor. Take the case of autism, for example. Affected people have a problem with theory of mind and communication, and they can be seriously challenged in the strictly linguistic domain as well (Fisher and Marcus 2006). The prediction is that there will be several-to-many genes found that will have pleiotropic effects on more than one spoke of the wheel in Figure 2.3.

Penn and Povinelli (2007) review evidence in favor of the idea that apes have no understanding of cause and effect in the physical domain. Wolpert (2003) points out that this seriously limits apes’ capacity to make and use tools; he even suggests that understanding causality may have given us language. But why are we so good at causal inference? Penn and Povinelli (2007, p. 111) suggest that it is our capacity of analogical reasoning which helps us to figure out more causes:

It is well known that human subjects often learn about novel and unobservable causal relations by analogy to known and/or observable ones: The structure of the atom, for example, is often described by analogy to the solar system; electricity is conceived of as analogous to a flowing liquid; gravity is like a physical force.

It is here where we see a possible close connection to language. First, many grammatical constructions function such that if one element appears, it causes
the mandatory appearance of another. Second, a sentence like *Mary loves John* is analogous to *Susan loves Jim*: language rests on an unlimited variety of such analogical constructions. So, besides recursion, analogy seems to be a strong connection between tool making and language; incidentally, it links strongly to causal inference as well. Regarding the sequential order of these adaptations in evolution we stress again that it may be an ill-formulated question: as stated at the onset of this chapter, it is more likely that the target of selection shifted among components of the human adaptive suite, and that improvement in one component may have given advantage in some others as well. In fact, any genetic change providing pleiotropic advantage in more than one domain is more likely to become fixed than one with the same positive effect in only one domain.

Analogical reasoning is likely to be an intermediate phenotype that is necessary for causal reasoning, tool making, and language. The same may hold for shared intentionality (Tomasello and Carpenter 2007): analysis of child development suggests that gaze following develops into joint attention, social manipulation transforms into cooperative communication, group activity develops into collaboration, and social learning develops into instructed learning. A third such intermediate skill is recursive processing in intentionality, tool making, and language. Fine motor control (Calvin 1991) is an executive skill that is required for the efficient implementation of human adaptations.

It is apparent that components of the human-specific adaptive suite can be tentatively grouped into two categories: (a) indispensable procedural
components (handling hierarchies, analogical reasoning, imitation, shared attention and intentionality, and fine motor control) and (b) complex adaptive faculties (docility, complex cooperation and theory of mind, language, and tool making).

This proposal raises a question about the evolution of such a complex network of procedures and interactions. As stated, pleiotropy is likely to be the rule here. If so, a concern can be raised about the evolutionary plausibility of such a suite, given the fact that many think that pleiotropy retards evolution (Orr 2000). Griswold (2006) has approached this question from a different perspective. First, he demonstrated that if a mutation with pleiotropic effects is overall beneficial, then it is likely to be beneficial for more than one trait. Second, such a mutation, although rare, will spread faster and have a higher rate of fixation in the population than others. Third, the rate of evolution of a phenotypic character may not decline when that character is pleiotropically associated to an increasing number of other characters, provided that the characters are under pure directional selection such that they are far from their optima relative to the average magnitude of a mutation. Griswold notes that such a situation is typical for adaptive radiations. We think that in the past 5 million years, such a radiation has happened in the hominine lineage.

Selective Scenarios for the Origin of Language

The origin of human language has provided fertile grounds for speculation and alternative theories have been proposed (Table 2.1).

Most of the theories that suggest a given context for the evolution of human language attempt to account for its functional role. Given that all of these theories are functionally more or less plausible, it is almost impossible to decide on their usefulness based only on this criterion. However, recent game theoretical research can help us evaluate various contexts. These criteria concern the interest of communicating parties and the cost of equilibrium signals.

The central issue is whether early linguistic communication was honest. If signal cost is the same for all signallers, then honest cost-free signalling can be evolutionarily stable only if there is no conflict of interest between the participants (Maynard Smith 1991). If the cost of signals varies with the quality of the signaller, then the situation is more complicated. In this case, it is possible to construct cost functions that give an arbitrarily low cost at equilibrium even if there is a conflict of interest (Hurd 1995; Számadó 1999; Lachmann et al. 2001) (Table 2.2). In the case of human language, the most obvious way to construct such a cost function is to punish dishonest signallers (Lachmann et al. 2001). This solution assumes, however, that dishonest signallers can, on average, be detected (i.e., signals can be cross-checked); it also assumes that dishonest signallers are punished (which is a nontrivial assumption). Thus, one can conclude that “conventional” signals will be used when communicating about (a) coincident interest or (b) verifiable aspects of conflicting interest;
### Table 2.1 Alternative theories to explain language evolution.

<table>
<thead>
<tr>
<th>Theory (Source)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gossip (Power 1998)</td>
<td>Menstrual ritual can be a costly signal of commitment; hence participating in such rituals can create female groups of shared interest in which sharing information about the social life of others (i.e., gossiping) can be beneficial</td>
</tr>
<tr>
<td>Grooming (Dunbar 1998)</td>
<td>Language evolved as a substitution for physical grooming. The need for this substitution derived from the increasing size of the early hominid groups.</td>
</tr>
<tr>
<td>Group bonding and/or ritual (Knight 1998)</td>
<td>Language evolved in the context of intergroup rituals, which first occurred as a kind of “strike action” against non-provisioning males. Once such rituals were established, a “safe” environment was created for further language evolution</td>
</tr>
<tr>
<td>Hunting (Washburn and Lancaster 1968); (Hewes 1973)</td>
<td>Our intellect, interests, emotions, and basic social life are evolutionary products of the success of the hunting adaptation. Later, Hewes (1973) argued that the probable first use of language was to coordinate the hunting effort of the group.</td>
</tr>
<tr>
<td>Language as a mental tool (Burling 1993)</td>
<td>Language evolved primarily for the function of thinking and was only later co-opted for the purpose of communication</td>
</tr>
<tr>
<td>Mating contract and/or pair bonding (Deacon 1997)</td>
<td>Increasing size of the early hominid groups and the need for male provisioning also necessitated “social contract” between males and females</td>
</tr>
<tr>
<td>Motherese (Falk 2004)</td>
<td>Language evolved in the context of mother–child communication: Mothers had to set their babies down to collect food efficiently, and their only option to calm them was to use some form of vocal communication</td>
</tr>
<tr>
<td>Sexual selection (Miller 2001)</td>
<td>Language is a costly ornament that enables females to assess the fitness of a male. According to this theory, language is more elaborate than a pure survival function would require.</td>
</tr>
<tr>
<td>Song (Vaneechoutte and Skoyles 1998)</td>
<td>Language evolved rapidly and only recently through cultural evolution, assuming two important sets of preadaptations: the ability to sing and better representation abilities (i.e., thinking and mental syntax).</td>
</tr>
<tr>
<td>Status for information (Desalles 1998)</td>
<td>Language evolved in the context of a so-called “asymmetric cooperation,” where information beneficial to the group was traded for status.</td>
</tr>
<tr>
<td>Tool making (Greenfield 1991)</td>
<td>Assumes a double homology: a homologous neural substrate for early ontogeny of the hierarchical organizations shared by two domains—language and manual object combination—and a homologous neural substrate and behavioral organization shared by human and nonhuman primates in phylogeny.</td>
</tr>
</tbody>
</table>
Table 2.2 To evaluate the properties and the explanatory power of the various theories, the following questions were asked: (1) Can the theory account for the honesty of early language, i.e., is there a shared interest between the proposed communicating parties? (2) Are the concepts proposed by the theory grounded in reality? (3) Can the theory account for the power of generalization unique to human language? (4) Can the theory account for the uniqueness of human language? T: thought; V: vocalization; G: gestures.

<table>
<thead>
<tr>
<th>Theory (Source)</th>
<th>Modality</th>
<th>First words</th>
<th>Topic</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gossip (Power 1998)</td>
<td>V</td>
<td>“Faithful,” “philander”</td>
<td>Social life</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Grooming (Dunbar 1998)</td>
<td>V</td>
<td>?</td>
<td>?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Group bonding and/or ritual (Knight 1998)</td>
<td>?/V</td>
<td>?</td>
<td>?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hunting (Washburn and Lancaster 1968); (Hewes 1973)</td>
<td>G/V</td>
<td>Prey animals</td>
<td>Coordination of the hunt</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Language as a mental tool (Burling 1993)</td>
<td>T</td>
<td>?</td>
<td>?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Mating contract and/or pair bonding (Deacon 1997)</td>
<td>?</td>
<td>?</td>
<td>Social contract</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Motherese (Falk 2004)</td>
<td>V</td>
<td>“Mama”</td>
<td>Contact call</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sexual selection (Miller 2001)</td>
<td>?</td>
<td>?</td>
<td>Anything</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Song (Vaneechoutte and Skoyles 1998)</td>
<td>V</td>
<td>?</td>
<td>?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Status for information (Desalles 1998)</td>
<td>?</td>
<td>?</td>
<td>Valuable information</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Tool making (Greenfield 1991)</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

“costly” signals will be used otherwise (Lachmann et al. 2001). Although theory thus far says nothing about the evolution of such systems of communication, there are a few computer simulations which suggest that honest, cost-free communication evolves only if there is shared interest between the participants (Bullock 1998; Noble 2000; Harris and Bullock 2002).

What does this tell us about the emergence of human language? The production cost of speech or gesturing appears to be low; thus human language consists of cost-free or low-cost signals at equilibrium (not counting time constraints). Based on the above criteria, one should favor either those theories which propose a context with no conflict of interest (e.g., hunting, tool making, motherese, grooming or group bonding, and/or ritual theory) or a context in which there might be a conflict of interest but where signals can be easily
cross-checked. None of the theories fit the second context. For example, both mating contract and gossiping assume a context in which conflict of interest exists and signals cannot be easily cross-checked.

Explaining the evolution of human language is likely to remain a challenge for the coming decade. Presently, no single theory is able to answer sufficiently all the questions about honesty and groundedness, power of generalization, and uniqueness. Table 2.2 gives a summary of these criteria (Számadó and Szathmáry, submitted). As one can see, most of the theories fail to answer the majority of the questions. Perhaps the easiest criterion to fulfil is shared interest, as there are a number of social situations which assume shared interest between communicating parties (e.g., hunting or contact calls). Only two theories—tool making (Greenfield 1991) and hunting (Washburn and Lancaster 1968)—do significantly better than the others, as they can answer three out of the four questions asked of them (Table 2.2). Thus, it might be tempting to say that some combination of the two could provide a series of selective scenarios that would fit all of our criteria. The most notable conclusion, however, is that all theories fail to explain the uniqueness of human language. Thus, even though indirect evidence strongly suggests that the evolution of human language was limited by selection, it remains difficult to envisage a scenario that would explain why.

Although the different scenarios suggest all kinds of selective forces, none of these scenarios has been consistently implemented in a family of models. Given the limitations of experimentation on humans and chimps, researchers should consider implementing the different scenarios in various model-based settings. Ultimately, researchers should be able to reenact the emergence of language in artificial worlds, many of which will probably involve robots. The use of robots offers a unique and probably indispensable way of symbol grounding (basic words, via concepts, should be linked to physical reality; Steels 2003) and somatosensory feedback (actions, or results of actions, on behalf of the agent feed back into its own cognitive system via sensory channels; Nolfi and Floreano 2002).

Some major transitions in evolution (such as the origin of multicellular organisms or that of social animals) happened a number of times, whereas others (the origin of the genetic code, or language) seem to have been unique events (Maynard Smith and Szathmáry 1995). One must, however, be cautious with the word “unique.” Due to a lack of the “true” phylogeny of all extinct and extant organisms, one can give it only an operational definition (Szathmáry 2003). If all the extant and fossil species, which possess traits due to a particular transition, share a last common ancestor after that transition, then the transition is said to be unique. Obviously, it is quite possible that there have been independent “trials,” as it were, but we do not have comparative or fossil

---

1 One example is the ECAgents, a project sponsored by the Future and Emerging Technologies program of the European Community; see http://ecagents.istc.cnr.it/).
evidence for them. What factors, then, can lead to “true” uniqueness of a transition? (a) The transition is variation-limited. This means that the set of requisite genetic alterations has a very low probability. “Constraints” operate here in a broad sense. (b) The transition is selection-limited. This means that there is something special in the selective environment that can favor the fixation of otherwise not really rare variants. Abiotic and biotic factors can both contribute to this limitation. For example (Maynard Smith 1998), a single mutation in the hemoglobin gene can confer on the coded protein a greater affinity for oxygen: such a mutation got fixed in some animals which live at high altitudes only (such as the lama or the barred goose, the latter migrating over the Himalayas at an altitude of 9000 m).

There are interesting subcases for both types of limitation. For (a), one can always enquire about the time-scale. “Not enough time” means that given a short evolutionary time horizon, the requisite variations have a very low probability indeed, but this could change with a widened horizon. An interesting subcase of (b) is “pre-emption,” meaning that the traits resulting from the transitions act via a selective overkill and sweep through the biota so quickly to competitively suppress further evolutionary trials. The genetic code could be a case in point.

Acknowledgments

The authors thank Mauro Santos, Luc Steels, Michael Corballis, Peter Gärdenfors, and Chrisantha Fernando for useful discussions. Partial support of this work has been generously provided by the National Office for Research and Technology (NAP 2005/KCKHA005) and the Hungarian Scientific Research Fund (OTKA NK73047).