

60 The Evolution of Language

W. TECUMSEH FITCH

ABSTRACT Progress in understanding the evolution of language requires a highly interdisciplinary integration of a vast amount of data from many fields. I present a framework for integrating such data from a comparative biological perspective, based on Tinbergen's four "why questions": mechanism, ontogeny, phylogeny, and function. I discuss three human capacities widely agreed to be necessary for language, but not present in chimpanzees: vocal imitation (speech), semantic reference, and recursive syntax. Although the combination of these abilities appears to be unique to humans, comparative data provide insights into each of them considered separately. Converging data from neuroscience, molecular genetics, linguistics, evolutionary theory and ethology are placing ever-tighter constraints on phylogenetic and functional hypotheses about language evolution. I argue that integration of these data is both possible and necessary and that a deeper understanding of language evolution is within our grasp.

Any attempt to take language evolution seriously entails an integration of neo-Darwinian evolutionary theory, speech science, and modern theoretical linguistics. In this review, I present an integrative empirical framework that emphasizes nascent areas of agreement without neglecting unresolved controversies. I will adopt the pluralistic approach to biological questions developed by Tinbergen (1963) in ethology and widely accepted by contemporary evolutionary biologists. Tinbergen pointed out that there are several equally valid, complementary approaches to answering biological questions, which can be placed into four categories: mechanistic, ontogenetic, phylogenetic, and functional. *Mechanistic* questions address the physiological or neural mechanisms that underlie individual behavior, and *ontogenetic* questions concern the developmental processes by which these individual mechanisms are constructed. *Functional* questions address the selective value of the mechanisms: Why did (or do) ancestral organisms possessing certain mechanisms outreproduce conspecifics lacking them? *Phylogenetic* questions concern the evolutionary history of a mechanism or species. The answers to each of these four questions influence and constrain the others, and a full understanding of any biological trait demands rich and mutually consistent answers to all of them. This is as true of biolinguistics (the biological study of language and its evolution (Lenneberg, 1967; Lieberman, 1984; Jenkins, 1999) as of any other biological enterprise. Thus, although any individual researcher will typically focus on one or two questions, the field as a whole must encourage, and attempt to integrate, research on all of them.

W. TECUMSEH FITCH School of Psychology, University of St. Andrews, St. Andrews, Fife, Scotland

Biolinguistics: A comparative evolutionary approach

The comparative method—the use of comparisons among species to derive inferences about phylogenesis and function—is a pillar of evolutionary biology. Evolution occurs slowly, with the demoralizing consequence that laboratory experiments on evolution are at best tediously slow and more typically impossible. Fortunately, Darwin recognized that each living species represents an “experiment of nature,” and by judiciously comparing extant species we often find that answers to important questions were provided long ago by evolution. The incorporation of comparative data by molecular biologists has led to stunning insights into the nature of genetic systems and the evolution of life (Carroll, Grenier, and Weatherbee, 2001), and comparative neuroscience has vastly enriched our understanding of neural mechanisms and their evolution (Allman, 1999). Although application of the comparative method to biolinguistics is just beginning, early results look promising (Hauser, Chomsky, and Fitch, 2002), and I will adopt this comparative approach here whenever possible.

A major conceptual hurdle in adopting the comparative approach to biolinguistics is the lack of language in extant nonhuman species: language is so different from the communication systems known in other animals that the very comparison may seem forced. Thus, we need to construct a model of language evolution that explains the undeniable differences between language and other communication systems in a manner that neither neglects the similarities nor trivializes the differences. Ultimately, a mature biolinguistics should be able to single out a set of crucial genetic changes, explain their effects in terms of brain development and neural function, relate these physiological mechanisms to the cognitive and computational principles of language, and understand why these particular changes were adaptive during our species's evolution. Although the answers to such questions seemed hopelessly distant a few decades ago, tantalizing findings in molecular genetics and the rapid progress that has been made in neurolinguistics and neural development offer grounds for cautious optimism, and substantial progress can be expected in the coming decades.

Defining the target: The faculty of language

Following Saussure (1916) and many others, I will refer to the system that underlies the human language capacity as

the *faculty of language*, or FL. Thanks to FL, any normal human child given adequate exposure to a heterogeneous set of linguistic utterances will develop a linguistic ability that not only encompasses these past utterances but also provides a limitless capacity to generate and understand novel utterances in the same style, providing a remarkable capacity to share novel concepts between individual brains. The fact that each of us went through this acquisition process, with little fanfare or conscious effort, should not blind us to its complex nature. The language faculty is responsible for the enormous cognitive and cultural power of our species, for good or ill. An infant chimpanzee raised under the same circumstances and exposed to the same corpus of utterances will not develop this capacity. This logically entails the recognition of a species-specific biological component of human language, which I will denote "FL in the narrow sense" (FLN), making no assumptions about its particular nature. It should be noted that the existence of an FLN distinguishing humans from chimpanzees is a truism and by no means licenses disregarding the data on what a chimpanzee *does* acquire in this situation. Indeed, such data provide one of the richest sources of comparative evidence about what is and is not uniquely human (Savage-Rumbaugh et al., 1993). Such comparative data paint an increasingly rich picture of what humans share with other apes, other mammals, and other vertebrates, and of the important role the shared components play in human language abilities. Thus FL, in the broad and inclusive sense (FLB), includes a wide variety of subsystems that play a crucial role in language but are shared with other animals, and a small subset of systems, capacities, or propensities that are unique to our species (FLN). A core desideratum of biolinguistics is to identify the nature and status of all of these subsystems (Hauser, Chomsky, and Fitch, 2002).

The shared components of FLB include, uncontroversially, most aspects of the biological and neural functioning of the human organism: FLB rests on a broad, shared foundation of basic capabilities that we inherited from our prelinguistic ancestors. These components of FLB can be studied in other species to gain insight into their functions in humans. Comparative studies also provide the only grounds, logically speaking, for claims that some system or capacity is uniquely human.

Current data justify provisionally singling out at least three potential candidates for uniquely human components of FLN: imitation, semantic reference, and recursion. None of these capacities is definitional of FLN. They are simply hypotheses, open to debate and empirical test, about what might constitute the uniquely human component of the language capacity. For example, Alvin Liberman and his colleagues suggested that categorical perception of speech sounds was a component of FLN, but this hypothesis was rejected when categorical perception was discovered in chin-

chillas and monkeys (Morse and Snowdon, 1975; Kuhl and Miller, 1978). However, other aspects of speech perception and production might still be uniquely human but have not yet been studied in nonhuman animals. A topic of intense current interest concerns various aspects of social intelligence that comprise "theory of mind," and it seems possible that some of these are uniquely human components of FLN. Thus, the provisional list above is probably incomplete. It is intended simply as a starting point for discussion and empirical investigation.

IMITATION Imitation is a fundamental human characteristic that underlies all aspects of human culture, including ritual, dance, visual art, games, music, and technological development. There is a large literature on imitation that makes many important distinctions among different types (e.g., mimicry vs. role-reversal imitation; Tomasello, 1999). Here I use the term in its simplest form: imitation is the ability to reproduce a novel signal or gesture produced by another individual. Vocal imitation in particular is a critical component of spoken language, without which a flexible but shared vocabulary could never develop. Although extremely sophisticated forms of imitation may be uniquely human, imitation in general is not. Indeed, a wide variety of species possess imitative abilities in the vocal domain, including birds, seals, and cetaceans (dolphins and whales). However, a clear and surprising conclusion of modern ethological studies is that our nearest relatives, the nonhuman primates, have very poor imitative capacities in any domain, and virtually no ability for vocal imitation. Apes do not "ape," and "monkey see, monkey do" is a highly misleading myth. Thus, imitation provides a nice example of the value of the comparative approach: we can both identify a phylogenetic lacuna between humans and apes that was spanned during our recent evolution, and we can study more distant relatives, such as birds and dolphins, to gain insights into the ontogeny, phylogeny, function, and mechanisms underlying vocal imitation (Fitch, 2000).

SEMANTIC REFERENCE Semantic reference, which gives language meaning, is roughly the process by which streams of sound (or gestures, in signed language) are systematically associated with conceptual structures in a consistent, intentional, and socially shared manner (see Jackendoff, 2002, for further discussion). "*Intentional*" is used here in a technical sense (Dennett, 1983); in this context it simply means "designed for use in meaningful exchanges" and does not necessarily connote any conscious awareness or "intent." Bee dances are intentional (designed by evolution for communicating flower locations) but most likely are unconscious on the signaling bee's part. Reference and meaning are *sine qua non* for language, and distinguish music (another universal human behavior) from language. Reference presup-

poses several subcapabilities, including some very basic ones, such as the ability to associate sounds and events, that are broadly shared with other vertebrates. An interpretive component, an ability to interpret conspecific vocalizations, gestures, or body postures as indicating certain cognitive or emotional states in the signaler, is another widely shared subcapability underlying many animal communication systems. In some cases, as in alarm calling, the state indicated may be associated with an external event, such as the appearance of predators of various types (Cheney and Seyfarth, 1990), and such systems are often termed “functionally” referential to distinguish them from true reference. At least in birds and mammals, studies of “audience effects” further indicate the capability on the part of signalers to modify or inhibit their behavior, depending on the presence or type of conspecific (Evans and Marler, 1994). Finally, the honeybee dance language, while limited to locational information, seems to represent a simple form of reference (including displacement of the referent and interchangeability between signaler and receiver; see Hockett, 1960). Such similarities to reference in language have led some authors to posit, or even implicitly assume, that reference is a widely shared capacity, long predating language. This conclusion is unjustified, for two reasons.

First, in the animal communication systems just mentioned, the signalers emit vocalizations from a limited, species-specific, and genetically canalized repertoire, and thus the number of messages expressible is strictly, genetically, limited. This is not true of linguistic utterances, which occur in limitless variety. Second, linguistic reference entails a symmetry between signaler and receiver, such that both parties attribute nearly the same meaning to the same utterance. This does not appear to be the case in primate communication: although a vervet’s cry of fear versus surprise may enable a listener to infer the presence of a leopard versus a snake, current data do not support the further supposition that the caller is intentionally (either consciously or unconsciously) referring to snakes or leopards (Seyfarth and Cheney, 2003). But it is precisely this design for communicating specific referents, pragmatically taking the listener’s knowledge into account, and the propensity to interpret utterances as intentional in this way that give language its limitless extent and great power (Tomasello, 1999). Thus, despite the great value of comparative data on “functionally referential” animal communication systems, rich referentiality appears, on current knowledge, to be a unique capability of human language. Note that this is not true by definitional fiat, or tautologically. If we discovered tomorrow that the learned and indefinitely extensible songs of mockingbirds are consistently associated with conceptual states and are used intentionally to inform other mockingbirds of these states, we would conclude, by the definition above, that mockingbirds have reference.

RECURSION Recursion is the capability that lends to language its limitless extensibility, especially in the realm of syntax and semantics. In computer science, recursive functions are those that “call themselves,” taking their own past outputs as future inputs. Recursive functions are differentiated from the much broader category of iterative or “feedback” functions, whose past output simply influences their future behavior. Iterative functions are ubiquitous in both biology (e.g., in motor control) and engineering (e.g., thermostats or antilock brakes), whereas the computational apparatus underlying recursion seems to be considerably rarer in nature. Recursion allows relatively simple embedding operations to build indefinitely complex structures. In linguistic recursion the types of data that serve as input and output are complex and structured (usually depicted as treelike phrase structures), and the entire structure can be operated on repeatedly without being flattened out, blended, or averaged, or otherwise lost, as in the following:

Mary is rather witty.
 John thinks (Mary is rather witty).
 Susan said (John thinks (Mary is rather witty)).
 I’m shocked that (Susan said (John thinks (Mary is rather witty))).

Such embedding of phrases within phrases is the hallmark of linguistic recursion.

Second, despite the continuously variable nature of the output medium (whether sound or gesture), linguistic structures such as words or sentences are discrete. Combined with the capacity of all recursive functions to generate an infinity of output, this discretization of the sensory signal provides the “discrete infinity” capacity that is central to linguistic recursion. This allows us to generate a limitless variety of utterances that vary in arbitrary but perceptible ways from one another. This capacity is central to the limitless expressivity characteristic of human language, as in other unlimited systems, such as chemistry or genetics (Studdert-Kennedy, 1998). Although some ability to iteratively construct hierarchical structures appears (like imitation) to be shared with birds and whales, although without any known semantic content, there is no evidence for such abilities in nonhuman primates.

Summarizing, I argue that at least three key characteristics of language, imitation, reference, and recursion, combine in our species to create the FLN. Although each component invites further study of similar capacities in other animals, the combination appears to be unique to our species. With three important mechanisms underlying FLN now identified, I will analyze language evolution from the fourfold Tinbergian perspective, starting with function.

What selective advantages did our hominid ancestors obtain from the precursor mechanisms ancestral to those underlying the modern FL? Because this question is unabashedly adaptationist, and because some thinkers on language evolution have questioned the concept that language is an adaptation, it is important first to clarify this question. A founding insight of modern evolutionary theory is that adaptation is “an onerous concept” (Williams, 1966), to be invoked carefully and interrogated rigorously in each particular case. A wide variety of constraints prevent organisms from being perfectly adapted (the perfect organism would live forever and produce infinite offspring), and any organism possesses nonadaptive features. Of course, the fastest empirical way to discover subtle nonadaptive features is to provisionally assume adaptiveness, construct plausible adaptive hypotheses, and then reject them one by one. This methodological imperative to assign adaptation to the role of null hypothesis may account for the persistence of Panglossian caricatures of adaptationists but in no way indicates a neo-Darwinian conviction that all features are adaptations. Chomsky’s view on the subject, that language as a whole is adaptive, was made clear in 1975: “There is an obvious selective advantage in the ability to discover the language of one’s speech community” (Chomsky, 1975, p. 252). Misunderstanding appears to derive mainly from Chomsky’s use of the term *language* in a technical sense to refer to certain core aspects of syntax (e.g., recursion/discrete infinity). That language as a whole is adaptive is hardly questioned by scholars today (Pinker and Bloom, 1990, Häuser, Chomsky, and Fitch, 2002; Jackendoff, 2002). Of course, both the particular aspects of FL that are, or are not, adaptations, and the nature of the specific functional benefits they confer, are open empirical questions.

The selective context(s) in which language, and particularly semantic reference, derived its advantage is a topic that has received inadequate attention. There is a widespread assumption that sexual selection and increased mating success played a critical role in the conferring of a selective advantage on individual language users. For example, (“females would surely have preferred mates whose communicative capacities so strikingly outclassed those of other available partners”—Bickerton, 1998, p. 353) and “that tribal chiefs are often both gifted orators and highly polygynous is a splendid prod to any imagination that cannot conceive of how linguistic skills could make a Darwinian difference” (Pinker and Bloom, 1990, p. 725). Mating success has also been invoked in the opposite context: “Subjacency has many virtues, but I am sure that it could not have increased the chances of having fruitful sex” (Lightfoot, 1991, p. 69). Despite its apparent appeal among academics, I know of no data supporting the notion that “better lan-

guage leads to more sex,” either in modern or in traditional societies. Nonetheless, the sexual selection hypothesis for language evolution appears to be the default assumption by many if not most scholars (Miller, 2001).

An alternative, the “mother tongues” hypothesis, is that the adaptive value of language lies primarily in enhancing communication among kin, particularly between parents and offspring (Fitch, in press). According to this argument, kin selection provides a viable evolutionary route to the capacity to honestly communicate arbitrary information. Young primates have a long childhood during which they are cared for by, and learn from, their mother and other related individuals, and any communicative efficiencies that enhance this learning process would offer a selective advantage in terms of offspring survival. This period of dependence is particularly long in humans and chimpanzees, and thus by inference in our prelinguistic ancestors, making offspring survival of paramount importance in human and ape reproductive success. Two facts support kin selection rather than sexual selection as the driving force behind language evolution. First and most obvious is the remarkable precocity of language development. Sexually selected traits typically develop at puberty, but a 1-year-old child is already far along in the process of language acquisition, and language is fully functional in humans many years prior to sexual maturity. Second is the lack of strong sexual dimorphism in language abilities: sexually selected traits are typically more highly developed in males, but language is sexually egalitarian. What small differences exist are in favor of females, who have larger vocabularies, earlier linguistic maturation, excel at tongue-twisters, and are less likely than males to suffer congenital language disturbances (Henton, 1992). Both factors are compatible with the mother tongues hypothesis, which would select for early linguistic competence and female bias and argue against sexual selection for semantic reference. This topic deserves further investigation.

A second functional issue concerns internal versus external uses of language. Is (or was) the primary functional advantage of early language capabilities for private thought within an individual or for interindividual communication? Clearly, both uses provide benefits in modern humans. The former hypothesis has been favored by some authors (e.g., Bierwisch, 2001), largely because it neatly solves a logical problem: with whom would an individual with some new advantageous mutation speak? However, this problem disappears if the primary beneficiaries of communicative acts are one’s own offspring, who are likely to share the mutation. Two further considerations blunt the internalist hypothesis further. First, while the ability to manipulate semantic conceptual structures internally is clearly of substantial value, it is unclear why such manipulation should occur in terms of externalizable words and phrases. The medium of “silent speech” is not a pure language of thought

but manifests itself in the phonological forms of our speech community, complete with rhyme, alliteration, and other semantically superfluous traits. The phonological component is entailed by communication, but seems without value from the internalist perspective. Second, the value of internal language use derives in large part from its association with externally derived knowledge. It is precisely the value of this shared cultural knowledge that makes our internal dialogue so meaningful and useful (Tomasello, 1999). True, an ability to internally and idiosyncratically name objects and events for personal reference might have some mnemonic value. But the value of this is dwarfed by the vastly increased power of linguistic thought enabled by communication between individuals, and the cultural accretion of knowledge that results. Thus, although the internalist hypothesis might provide a phylogenetic starting point for certain aspects of FL, such as recursion, it cannot explain FL as a whole, and particularly not the phonological component that renders language so well-suited to communication between individuals and to the cultural transmission of knowledge.

Ontogeny

The vast field of language acquisition contributes to linguistics in general and to biolinguistics in particular. The speed and robustness of language acquisition in childhood are among the chief reasons that many linguists adopt a nativist viewpoint. A child exposed to a hodgepodge of primary linguistic data, some of it incomplete or erroneous, will nonetheless reliably arrive at an abstract rule system that overlaps almost completely with that of other individuals in its community. The richness of the rule system, compared with the messiness and incompleteness of the input (the so-called "poverty of the stimulus"), has led many thinkers to conclude that at least some aspects of this rule system must be innate. This is a complex and extremely controversial topic that I will mostly avoid here (see Crain, 1991, and commentary for further discussion). However, from the biolinguistic perspective advocated here, two points are worth making. First, the questions of "learnability" that traditionally dominate discussions of this topic should not be considered in a vacuum but must be balanced against issues of implementability that take the nature of neural and genetic mechanisms into account, and against issues of evolvability that take the phylogenetic and functional history of our species into account. Second, irrespective of the possible existence of innate rules, there is a vast amount of linguistic structure that varies completely between languages and thus must be learned, namely, the contents of the lexicon. Despite the underspecified nature of the lexical input (a child typically has simply an utterance, unparsed into words, and some context, unparsed into meaning, to work with),

children do indubitably acquire a vast lexicon, very close to that of their community, without the aid of any innate lexical items. Lexical acquisition is thus an ability to which innate constraints on learning, rather than innate knowledge, must apply.

Research in lexical acquisition has highlighted the remarkable speed and accuracy with which children acquire word meanings and has uncovered a number of basic principles that children appear to use to constrain their inferences about meaning, such as the *mutual exclusivity assumption* (that each object receives a single label) or the *whole object assumption* (that words label entire objects) (Markman, 1990). This raises the possibility of innate constraints on lexical inference, which would be potential candidates for inclusion in the FLN. However, recent data (Markson and Bloom, 1997) suggest that these same principles apply equally to the learning of nonlinguistic facts. Thus, lexical acquisition may be a case in which innate cognitive constraints on social cognition and learning, independent of language and presumably with a longer evolutionary history, play a crucial role in language acquisition. If true, this is good news for biolinguists, for two reasons. First, it means we can use comparative data from nonhuman cognition to help understand this critical component of the language faculty, and second, it lightens the load that evolutionary explanations must bear by extending the origin of innate constraints much farther back into primate evolutionary history. The existence and importance of innate constraints on animal learning have been appreciated for many years (Garcia and Koelling, 1966), and to the extent that such shared constraints aid language acquisition, we have one less thing to explain in the recent evolution of our species. Lexical acquisition thus provides a nice example of how results of studies on ontogenetic, mechanistic, and evolutionary mechanisms can be both mutually informative and mutually constraining.

Phylogeny

Human ancestors diverged from chimpanzees 5–6 million years ago, a short time in evolutionary history (for instance, vertebrates diverged from other animals more than 500 million years ago, and mammals from birds more than 200 million years ago). During that time a suite of changes occurred, some of which left a fossil record. Fossils provide clear evidence of at least three "grades" of human evolution, namely, australopithecines, *Homo erectus*, and *Homo sapiens*. These phases are distinguished by their brain and body size, tools, and geographic distribution. The earliest change was bipedalism, which long preceded the brain expansion and increasingly sophisticated tool use that typify later stages of human evolution. Unfortunately, language leaves no fossils, and even peripheral aspects such as vocal tract anatomy are inadequately preserved in fossils for any

consensus to have emerged about their timing (Fitch, 2000). Based on inferences from archaeological data, many experts estimate that full language emerged relatively recently (e.g., in the last 100,000 years or later; Mellars, 1998), but we have very little hard evidence to date this or previous stages, and all we know for certain is that all of the various capacities underlying human language evolved sometime during the last 6 million years, since our divergence from chimpanzees. Thus, models of language phylogeny are relatively unconstrained by fossil data, and a consequent indulgence in phylogenetic storytelling has given this component of the biolinguistic enterprise a poor reputation. However, non-fossil comparative data, though often neglected, can also constrain phylogenetic hypotheses, and can provide opportunities for testing them.

To give an example, it has frequently been suggested that the descent of the human larynx (Lieberman, 1984) resulted directly from bipedalism, gravity inevitably lowering the larynx in an upright organism (Wind, 1983), and this idea is sometimes treated as a fact (e.g., Boë et al., 2002). If true, the descent of the larynx would have occurred early, in *Australopithecus*. But an examination of the comparative data serves to reject this hypothesis. All birds are bipedal, spending their lives upright, but no species has a descended larynx. Among mammals, kangaroos have a hopping bipedal locomotory system that exerts considerable gravitational force on their internal organs. Although kangaroos share with humans antigravity specializations that are not seen in non-bipedal mammals, such as a closed inguinal canal (Coveney et al., 2002), their larynx lies in the normal, undescended mammalian position. This is one case out of many where comparative data can reject superficially plausible phylogenetic hypotheses that have been repeated for decades.

Another source of constraints on phylogenetic hypotheses is logical. If one explicitly anatomizes language into sub-components, as I've done here, the simple requirement that the acquisition of each subcomponent had to serve some purpose (not necessarily identical to its current usage) puts rather stringent restrictions on the construction of plausible phylogenetic hypotheses. Although partial phylogenetic hypotheses are simple to generate (e.g., assuming that reference and imitation are already in place, and then discussing only recursion), constructing a plausible and complete chain of events stretching from our common ancestor with chimps to the present FL, consistent with all the available evidence, is much more challenging. By way of example, I will sketch a framework for language phylogenesis consistent with the data reviewed thus far.

Many theorists posit a distinct intermediate stage of language evolution in which the signaling medium, complexity, and selective advantage of the protolanguage differed from modern language. A popular dual-stage theory is the gestural origins theory, dating back to at least Condillac (Hewes,

1973). Another detailed dual-stage model is that of Donald (1998), who posits a long intermediate "mimetic" stage of human culture in which humans had a rich imitative culture that included dance, social ritual, music, and tool making but lacked language. Putting some of these ideas together, I will offer a three-stage model that appears compatible with the currently available fossil and comparative data.

The first hypothetical stage, characterized by australopithecines, was prelinguistic. Upright posture led to increased tool use and exploitation of new resources (especially protein-rich carcasses), and led to larger brains, but communication remained chimplike. The second, or "protolinguistic" stage, associated with *Homo erectus*, was associated with prolonged selection for imitation, and corresponds to Donald's mimetic stage. Progressively greater imitative abilities enhanced social cohesion and cultural transmission, giving early *Homo* a huge advantage over other hominids. These were the first hominids to expand out of Africa, occupying most of the Old World. Although nonlinguistic, these mimetic hominids had a rich and extensible vocal repertoire (see also MacLarnon and Hewitt, 1999), based on imitation and surface-level recursion, comparable to that of mockingbirds or whales. However, these "songs" were only holistically or ritualistically related to meaning and were probably not referential or intentionally meaningful in a modern sense. Following Darwin's (1871) suggestion that early stages of language evolution were musical and used for courtship, a key role for such performances may have been mate choice, and thus driven by sexual selection. If so, these protolinguistic abilities would have been more highly developed in males. In the final, linguistic stage, in later *Homo*, the crucial transition to referential language was made, perhaps via an analytical "insight" that individual components of songlike utterances could be associated with individual components of concepts (e.g., objects vs. actions). In this last stage, kin communication played a crucial role, and the accurate communication of propositional information became the primary selective force driving further evolution to our current state (Fitch, in press). This hypothesis implies that we should seek "fossils" of protolanguage not in one-word utterances (Bickerton, 1998) but in complex music and song. Plausibly enough, music is a human universal, of uncertain selective value, precisely as we would expect of a holdover from an earlier evolutionary period. This music-as-protolanguage hypothesis highlights the value of comparing the formal structure and the neural basis of music and language (Lerdahl and Jackendoff, 1983; Zatorre, Belin, and Penhune, 2002).

Mechanism

The mechanisms underlying FL are a major focus of research in cognitive neuroscience. Earlier chapters in this

section dealt with neural mechanisms underlying language, and the evolution of peripheral mechanisms involved in speech and phonology have been reviewed elsewhere in detail (MacNeilage, 1998; Fitch, 2000; Lieberman, 2000). An area of exciting recent progress is the isolation of a gene, *FOXP2*, that differs between humans and chimps and is linked to oromotor coordination. Increasing oromotor control in humans was a critical prerequisite for vocal imitation and thus speech, and these findings provide a new source of information on the nature and timing of the changes undergone in linguistic evolution. The importance of this gene was discovered in long-term studies (Vargha-Khadem et al., 1995) of a large family, KE, some members of which have inadequate oral motor control (incorrect early claims that their deficiency was exclusively linguistic [Gopnik, 1990] were based on incomplete data). By isolating first the chromosomal location and finally the gene underlying this problem, these researchers opened the door to a comparison of the gene in humans and great apes (Enard et al., 2002). This comparison revealed small but crucial differences in the gene (only two amino acids differ) for the *FOXP2* protein, which regulates the expression of other genes. That small structural differences could have major phenotypic effects is exciting, because these are precisely the sorts of changes we must uncover to understand the substantial difference between humans and chimps in the context of nearly identical DNA. It is also intriguing that *FOXP2* is a highly conserved protein that shows little functional variation among the other mammals examined. These very recent advances are a promising example of the power of a highly interdisciplinary biolinguistic approach, and we can await experiments in transgenic mice in which the precise effects of the human version of the gene on mammalian neural development and function can be examined.

An important open mechanistic topic is the medium of language output. Significant progress in the study of signed language (Klima and Bellugi, 1979; Bellugi, Poizner, and Klima, 1990) leaves little doubt that the visual/gestural mode for linguistic expression is perfectly capable of supporting the full richness of human linguistic communication. Further studies of Tadoma (a tactile method of linguistic interchange developed for deaf and blind children) suggest the same is true of tactile communication (Vivian, 1966). That alternative linguistic expressive modes are available and fully adequate reveals a fundamental fact about FL and its neural instantiation: human linguistic abilities transcend any particular sensorimotor channel. However, it is equally relevant that the audiovocal channel is adopted in all human societies with the requisite capacities (disregarding the recent cultural innovation of reading), making it possible that there has been specific selection on the human auditory system for increased speech efficiency, as posited by many speech

scientists (Lieberman, 1984; Liberman, 1996). An oft-stated advantage of speech is that it leaves the hands free, but signers communicate remarkably well while working with their hands, and in any case have a complementary advantage over speech during eating or drinking. Similarly, communication in total darkness and silently is possible with Tadoma. Such factors thus seem incomplete as explanations for the dominance of speech as the medium of modern language, suggesting that additional neural or cognitive factors are at work.

The communicative adequacy of sign has led many researchers, dating back to Condillac, to posit a gestural origin of language (Hewes, 1973): that in an earlier stage of evolution, humans communicated primarily via signed gestures and vocalization was ancillary. Gestural origin hypotheses derive some of their apparent, persistent appeal from the communicative adequacy of sign language and the putative similarity between chimp gestures and sign (though see Tomasello, 1999). Recently, the discovery of mirror neurons has reinvigorated interest in gestural origins (Rizzolatti and Arbib, 1998; Corballis, 2002) and in the origins of imitation. However, the well-documented ineptitude of monkeys for manual imitation blocks any facile equation of mirror neurons and imitation skills, though mirror neurons perhaps were a "preadaptation" for imitation and thus language. In any case, mirror neurons would support gestural origin hypotheses only if corresponding audiovocal mirror neurons, which respond similarly to a private's own call and that of another private, were absent (and this seems unlikely, Kohler et al., 2002). Finally, gestural origin theories do not explain the genesis of the speech mode or its ascent to total dominance among modern humans. Backward extrapolation from modern signing capabilities to a primitive protolanguage thus seems both overly literal and unnecessary, given current data.

Conclusion: Taking language evolution seriously

One of the central insights of modern linguistics is the startling complexity of natural language. The fact that thousands of hard-working and highly intelligent linguists and engineers have yet to implement a full grammar of any language, despite many decades of concerted effort, gives some insight into the magnitude of the problem, which is masked by the effortlessness with which children acquire language. A central question for biolinguistics, then, is where this complexity comes from. Early approaches suggested that much of it was innately specified, in terms of a large set of language-specific rules called Universal Grammar (Chomsky, 1965), as argued by many scholars (Pinker and Bloom, 1990; Pinker, 1994; Jackendoff, 2002). However, there is increasing interest within linguistics in an alternative conception, broadly labeled the Minimalist Program

(Chomsky, 1994). By this hypothesis, the syntactic component of language is simple but powerful, composed of two operators, Move and Merge, and the vast complexity and variability of natural languages are derived from the phonological output component and (in particular) the semantic or conceptual component. This conception is appealing from the biolinguistic perspective, for several reasons. It lightens the load for evolutionary explanations: that a complex, interconnected suite of genetically specified rules specific to language could evolve to fixation in all human populations, in the short time since our divergence with chimps, is far less likely than the evolution of a few specific but powerful new capabilities. Most of the complexity underlying Universal Grammar would then derive from cognitive constraints or innate conceptual primitives that we share with other animals. Processing constraints associated with the phonological interface, and computational constraints associated with bridging the phonological and semantic components, would drive additional complexities. All of these factors can be studied via the comparative method, employing the full suite of modern neuroscientific techniques along with computer models and traditional cognitive experimentation. There is also a methodological advantage to assuming FLN is simple but FLB complex and shared with other animals. Empirical evaluation of the traditional assumption of a complex, language-specific Universal Grammar awaits a future synthesis of genetics, neural development, neuroscience, and linguistics that seems dauntingly distant. In contrast, the minimalist perspective generates hypotheses that can be empirically tested today, and such hypotheses thus provide an effective empirical strategy for discovering what is, and what is not, part of FLN.

ACKNOWLEDGEMENTS I thank Alfonso Caramazza, John Hyman and David Raubenheimer for comments on the manuscript.

REFERENCES

- ALLMAN, J. M., 1999. *Evolving Brains*. New York: Scientific American Library.
- BELLUGI, U., H. POIZNER, and E. S. KLIMA, 1990. Mapping brain function for language: Evidence for sign language. In *Signal and Sense*, G. M. Edelman, W. E. Gall, and W. M. Cowan, eds. New York: Wiley, pp. 521–543.
- BICKERTON, D., 1998. Catastrophic evolution: The case for a single step from protolanguage to full human language. In *Approaches to the Evolution of Language*, J. R. Hurford, M. Studdert-Kennedy, and C. Knight, eds. New York: Cambridge University Press, pp. 341–358.
- BIERWISCH, M., 2001. The apparent paradox of language evolution: Can Universal Grammar be explained by adaptive selection? In *New Essays on the Origin of Language*, J. Traban and S. Ward, eds. Berlin: Mouton de Gruyter, pp. 55–81.
- BOË, L.-J., J.-L. HEIM, K. HONDA, and S. Maeda, 2002. The potential Neandertal vowel space was as large as that of modern humans. *J. Phonet.* 30:465–484.
- CARROLL, S. B., J. K. GRENIER, and S. D. WEATHERBEE, 2001. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden, Mass.: Blackwell Science.
- CHENEY, D. L., and R. M. SEYFARTH, 1990. *How Monkeys See the World: Inside the Mind of Another Species*. Chicago: University of Chicago Press.
- CHOMSKY, N., 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.
- CHOMSKY, N., 1975. *Reflections on Language*. New York: Pantheon.
- CHOMSKY, N., 1994. *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- CORBALLIS, M. C., 2002. *From Hand to Mouth: The Origins of Language*. Princeton, N.J.: Princeton University Press.
- COVENEY, D., G. SHAW, J. M. HUTSON, and M. B. RENFREE, 2002. The development of the gubernaculum and inguinal closure in the marsupial *Macropus eugenii*. *J. Anat. (Lond.)* 201:239–256.
- CRAIN, S., 1991. Language acquisition in the absence of experience. *Behav. Brain Sci.* 14:597–650.
- DARWIN, C., 1871. *The Descent of Man and Selection in Relation to Sex*. London: John Murray.
- Dennett, D. C., 1983. Intentional systems in cognitive ethology: The “Panglossian paradigm” defended. *Behav. Brain Sci.* 6: 343–390.
- DONALD, M., 1998. Mimesis and the Executive Suite: Missing links in language evolution. In *Approaches to the Evolution of Language*, J. R. Hurford, M. Studdert-Kennedy, and C. Knight, eds. New York: Cambridge University Press, pp. 44–67.
- ENARD, W., M. PRZEWORSKI, S. E. FISHER, C. S. L. LAI, V. WIEBE, T. Kitano, A. P. MONAGO, and S. PAÄBO, 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418:869–872.
- EVANS, C. S., and P. MARLER, 1994. Food-calling and audience effects in male chickens, *Gallus gallus*: Their relationships to food availability, courtship and social facilitation. *Anim. Behav.* 47:1159–1170.
- FITCH, W. T., 2000. The evolution of speech: A comparative review. *Trends Cogn. Sci.* 4:258–267.
- FITCH, W. T. (in press). Evolving honest communication systems: Kin selection and “mother tongues.” In *The Evolution of Communication Systems: A Comparative Approach*, D. K. Oller and U. Griebel, eds. Cambridge, Mass.: MIT Press.
- GARCIA, J., and R. A. KOELLING, 1966. Relation of cue to consequences in avoidance learning. *Psychonom. Sci.* 4:123–124.
- GOPNIK, M., 1990. Feature-blind grammar and dysphasia. *Nature* 344:344–346.
- HAUSER, M., N. CHOMSKY, and W. T. FITCH, 2002. The language faculty: What is it, who has it, and how did it evolve? *Science* 298:1569–1579.
- HENTON, C., 1992. The abnormality of male speech. In *New Departures in Linguistics*, G. Wolf, ed. New York: Garland Publishing.
- HEWES, G. W., 1973. Primate communication and the gestural origin of language. *Curr. Anthropol.* 14:5–24.
- HOCKETT, C. F., 1960. Logical considerations in the study of animal communication. In *Animal Sounds and Communication*, W. E. Lanyon and W. N. Tavolga, eds. Washington, D.C.: American Institute of Biological Sciences.
- JACKENDOFF, R., 2002. *Foundations of Language*. New York: Oxford University Press.
- JENKINS, L. 1999. *Biolinguistics: Exploring the Biology of Language*. New York: Cambridge University Press.
- KLIMA, E. S., and U. Bellugi, 1979. *The Signs of Language*. Cambridge, Mass.: Harvard University Press.

- KOHLER, E., C. KEYSERS, M. A. UMILTÀ, L. FOGASSI, V. GALLESE, and G. RIZZOLATTI, 2002. Hearing sounds, understanding actions: Action representation in mirror neurons. *Science* 297:846–849.
- KUHL, P., and J. D. MILLER, 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *J. Acoust. Soc. Am.* 63:905–917.
- LENNEBERG, E. H., 1967. *Biological Foundations of Language*. New York: Wiley.
- LERDAHL, F., and R. JACKENDOFF, 1983. *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.
- LIBERMAN, A. M., 1996. *Speech: A Special Code*. Cambridge, Mass.: MIT Press.
- LIEBERMAN, P., 1984. *The Biology and Evolution of Language*. Cambridge, Mass.: Harvard University Press.
- LIEBERMAN, P., 2000. *Human Language and Our Reptilian brain: The Subcortical Bases of Speech, Syntax and Thought*. Cambridge, Mass.: Harvard University Press.
- LIGHTFOOT, D., 1991. Subjacency and sex. *Lang. Commun.* 11:67–69.
- MACLARNON, A., and G. HEWITT, 1999. The evolution of human speech: The role of enhanced breathing control. *Am. J. Phys. Anthropol.* 109:341–363.
- MACNEILAGE, P. F., 1998. The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21:499–546.
- MARKMAN, E. M., 1990. Constraints children place on word meanings. *Cogn. Sci.* 14:57–77.
- MARKSON, L., and P. Bloom, 1997. Evidence against a dedicated system for word learning in children. *Nature* 385:813–815.
- MELLARS, P., 1998. Neanderthals, modern humans and the archaeological evidence for language. In *The Origin and Diversification of Language*, N. G. Jablonski and L. C. Aiello, eds. San Francisco: Academy of Sciences.
- MILLER, G. F., 2001. *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. New York: Doubleday.
- MORSE, P. A., and C. T. SNOWDON, 1975. An investigation of categorical speech discrimination by rhesus monkeys. *Percept. Psychophys.* 19:137–143.
- PINKER, S., 1994. *The Language Instinct*. New York: William Morrow.
- PINKER, S., and P. BLOOM, 1990. Natural language and natural selection. *Behav. Brain Sci.* 13:707–784.
- RIZZOLATTI, G., and M. A. ARBIB, 1998. Language within our grasp. *Trends Neurosc.* 21:188–194.
- SAUSSURE, F. D., 1916. *Course in General Linguistics*. New York: McGraw-Hill.
- SAVAGE-RUMBAUGH, E. S., J. MURPHY, R. A. SEVCIK, K. E. BRAKKE, S. L. WILLIAMS, and D. M. RUMBAUGH, 1993. Language comprehension in ape and child. *Monogr. Soc. Res. Child Devel.* 58:1–221.
- SEYFARTH, R. M., and D. L. CHENEY, 2003. Signalers and receivers in animal communication. *Ann. Rev. Psychol.* 54:••–••.
- STUDDERT-KENNEDY, M., 1998. The particulate origins of language generativity: From syllable to gesture. In *Approaches to the Evolution of Language*, J. R. Hurford, M. Studdert-Kennedy, and C. Knight, eds. New York: Cambridge University Press, pp. 202–221.
- TINBERGEN, N., 1963. On aims and methods of ethology. *Z. Tierpsychol.* 20:410–433.
- TOMASELLO, M., 1999. *The Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard University Press.
- VARGHA-KHADEM, F., K. WATKINS, K. ALCOCK, P. FLETCHER, and R. PASSINGHAM, 1995. Praxic and nonverbal cognitive deficits in a large family with a genetically-transmitted speech and language disorder. *Proc. Nat. Acad. Sci. U.S.A.* 92:930–933.
- VIVIAN, R., 1966. The Tadoma method: A tactual approach to speech and speech reading. *Volta Rev.* 68:733–737.
- WILLIAMS, G. C., 1966. *Adaptation and Natural Selection*. Princeton, N.J.: Princeton University Press.
- WIND, J., 1983. Primate evolution and the emergence of speech. In *Glossogenetics: The Origin and Evolution of Language*, É. Grolier, ed. New York: Harwood Academic Publishers, pp. 15–35.
- ZATORRE, R. J., P. BELIN, and V. B. PENHUNE, 2002. Structure and function of auditory cortex: Music and speech. *Trends Cogn. Sci.* 6:37–46.