THE EVOLUTION OF LANGUAGE[1]

TECUMSEH W. FITCH

In 1866, the Linguistic Society of Paris issued a ban on all further discussion of the evolution of language. At the time, the ban seemed justified: the almost complete lack of relevant empirical data meant that the field was dominated by unfettered speculation, some of it bordering on fantasy. This threatened to undermine the reputation of the then-young field of linguistics. Unfortunately, however, the poor reputation of discussions of the evolution of language has continued to the present, with at least two negative consequences. First, the evolution of one of the most fascinating aspects of human biology, one which is arguably the core capability of our species, remains very poorly understood (indeed, David Premack called the evolution of language "an embarrassment for evolutionary theory", and his opinion seems widely shared). Second, the peripheralization of the field has seemed to provide a continuing license for unfettered speculation of the sort that led to the ban in the first place. The purpose of this talk is to argue that neither of these consequences is defensible today. The amount of empirical data relevant to the evolution of language is now considerable, flowing from a diverse set of disciplines including animal behavior, theoretical and comparative linguistics, neuroscience, anthropology, evolutionary theory, psychology, neuroscience, and bioacoustics. Indeed, this database is so vast and rapidly changing as to be beyond the mastery of a single individual. In particular, comparative data (data based on other living species) provides a rich but underutilized source of insights and ways to test evolutionary hypotheses. In fact, even a relatively cursory examination of the comparative data serves to call into question some of the most common assumptions about the evolution

[1] This talk was given on April 9, 2003 at the Wissenschaftskolleg zu Berlin as an evening colloquium. Because I gave the lecture spontaneously, without a written text, the text is an overview covering the main points I discussed. It was written several months afterward. This summary omits all of the empirical data presented, descriptions of the videos shown or sounds played, and all references.

of language. Thus, it seems that the time is ripe for the empirical study of the evolution of language to come of age.

I want to illustrate the value of the comparative approach by focusing first on the evolution of speech. Speech is an unusual capacity of our species, and no other primate has vocal capabilities (particularly in the realm of vocal imitation) that even approach those of a two-year-old child. The evolution of speech can and should be considered separately from that of language *per se*, because language is potentially expressible via many different sensory channels (the existence of writing, or signed languages, demonstrates this). Nonetheless, speech is used by all normal human communities as the primary vehicle for language and bears several indications of being a special adaptation. Because the speech signal is an objective, measurable entity and because speech production is well understood, our understanding of speech evolution is relatively advanced compared with other aspects of language evolution.

For many years it has been suspected that a crucial speech-related difference between humans and other animals was a descended larynx. In most mammals the larynx is located high in the throat and can be inserted into the nasal passages during breathing. Indeed, a newborn human infant starts life this way. This allows infants and many other mammals to breathe (through the nose) and swallow (through the mouth) simultaneously. However, early in life (starting at around three months), the human larynx begins to descend toward the lungs, pulling the tongue root with it. Although this presumably increases our risk of choking considerably (everything an adult human swallows must pass over the larynx), the descent of the larynx and tongue root also makes our vocal tract more versatile and allows us to produce speech sounds that would be impossible with a larynx in the normal mammalian position. Because several of these speech sounds are found in all human languages (e.g. the vowels in "beet" and "boot"), the descent of the larynx has been considered a unique human characteristic necessary for modern human speech and a major turning point in the evolution of our species.

But surprisingly, this conclusion was based on very little comparative data. Our prior knowledge of animal anatomy was derived from dead, formalin-fixed specimens that provide an inadequate indication of the physiology of the living animal. I thus set out to discover what animals do with their larynges during vocalization (which is what really matters, acoustically speaking). Working with A. W. Crompton at Harvard, I was able to film animals vocalizing (dogs, pigs, goats, and cotton-top tamarin monkeys). To our great surprise, although all of these animals do engage the larynx in the nasal cavity during quiet

breathing (as suggested by old anatomical studies), they lower the larynx during loud vocalizations, sometimes by a considerable distance. For example, during barking, dogs lower the larynx almost to their collar and thus for a moment attain a vocal anatomy quite similar to that of humans. These data suggest that temporary descent of the larynx during loud vocalization (presumably to increase the loudness of calls, which would be muffled if they passed through the nose) is a primitive mammalian characteristic, shared with most species.

However, the larynx in an adult human is *permanently* lowered, unlike in most mammal species. Thus, the permanently descended larynx is still arguably unique to our species. This idea also turns out, on examination of relevant comparative data, to be incorrect. Working with the French biologist David Reby, I was able to show that adult males in several deer species (red deer *Cervus elaphus* and fallow deer *Dama dama*) have a permanently descended larynx. A few years later, we now know that humans and deer are joined in this odd club by all the big cats (lions, tigers, leopards, and jaguars) and by koalas. Because none of these species use their descended larynx in the service of speech, I have proposed that the descended larynx in other animals serves the purpose of exaggerating the impression of size conveyed by vocalizations by lowering their formant frequencies. This hypothesis is grounded in considerable acoustic and anatomical data. Given that these same principles apply equally well to humans, it is possible that the original force driving the descent of the larynx during human evolution may had nothing to do with speech or language, but instead have simply served a size-exaggerating role, as for other animals. Only later in our evolution did we capitalize on this preadaptation to produce the wider range of phonemes that the descended larynx allows. This hypothesis is supported by the fact that the human larynx undergoes a second descent at puberty, but only in males. Given that this leads to no increase in the speech capabilities of teenage boys, the only plausible explanation is that this second laryngeal descent serves the same function in men as the descended larynx does in deer: to create an exaggerated impression of size (like the beard and broad shoulders developing at the same time).

Several conclusions can be drawn from this research. First, it is pointless to claim that a feature is "uniquely human" before a thorough search for the trait has been conducted throughout the animal kingdom. Though this is logically obvious, it is quite surprising how ready humans seem to be to conclude that some important human trait is unique to our species, even without looking at other species and without noting how often this mistake has occurred in anthropology and the cognitive sciences. Second, the existence of comparative data can provide a rich testing ground for hypotheses about evolution. If we

narrow our view to human evolution alone, any speculation seems reasonable, but once we include animal data (in effect examining the outcome of countless "experiments of nature", run over millions of years by evolution), the number of plausible hypotheses consistent with the data becomes quite small. Thus, our comparative database must be extended and considered seriously, if our understanding of the evolution of speech and language is to progress.

To highlight the promise of comparative research in understanding speech, I will discuss an important unsolved problem: the evolution of our ability to vocally imitate. Humans learn to produce new sounds based on the utterances of others and excel at this ability from a very early age (indeed we are all better as children than as adults). Vocal imitation is a crucial prerequisite for forming the large and open-ended vocabulary without which speech would be of limited use. Surprisingly, nonhuman primates have virtually no ability to imitate vocalizations (and indeed imitation skills in general are rudimentary in primates: "monkey see, monkey do" is a highly misleading bit of folk ethology). A chimp, raised in diapers alongside a human baby, will not spontaneously learn to say any words at all, and even with diligent reinforcement and teaching can produce only the poorest imitation of a few monosyllabic words. In sharp contrast, many bird species, dolphins and whales, and some seals are gifted vocal imitators. The abilities of mynahs or parrots to imitate speech are well-known. Less well-known is the story of Hoover, an orphaned harbor seal pup who was raised by a Maine fisherman and then donated to the New England Aquarium in Boston. At about five years of age, when he attained sexual maturity, Hoover started to produce speech, which not only was immediately recognizable as English but duplicated the dialect of his original caretaker. Research on birdsong has already catapulted songbirds to the forefront of progress in understanding vocal learning, and we now know that genes are being turned off and on and new neurons being born in the brains of birds as they learn their species-specific song. Thus, perhaps nonintuitively, the best place to look for a better understanding of the evolution of this crucial requirement of spoken language is not our nearest relatives (chimpanzees or other primates), but much more distantly related species such as dolphins, birds, or seals.

For the second part of this discussion, I will turn to the phylogeny of language, focusing on the two key properties of meaning and syntax. Through what process did these aspects of language evolve in our species? Although the very nature of this question might seem to render such discussion irredeemably speculative, I will argue that, even here, the comparative approach provides a crucial, and powerful, tool for evaluating evolutionary

hypotheses. It is not difficult to find comparative data that exclude statements commonly made in the speculative literature. For instance, it is often stated that the assumption of upright posture automatically led to a descended larynx, and this is sometimes quoted as if it were a self-evident fact. But the existence of many species that have upright posture but lack a descended larynx (including birds, kangaroos, and many arboreal primates) suffices to reject this hypothesis. Indeed, it seems that many of the hypotheses that are currently on offer in the literature can be tested and enriched by taking comparative data into account.

Take, as a crucial example, the evolution of propositional meaning in language. By "meaning" I mean our ability to intentionally transmit, honestly and informatively, arbitrarily complex thoughts from one individual's brain to another. The fact that we sometimes use language dishonestly must not obscure the fundamental fact that language can be used truthfully. This differentiates language from the vast majority of utterances that animals produce, particularly the most complex animal signals (like bird- or whale-song): as far as we can tell, these signals convey no propositional meaning beyond basic information (species, location, quality) about the signaler. How could the capability to convey propositions have evolved? What selective pressures could have favored such an ability? Many researchers have assumed that sexual selection could have sufficed, supposing that choosy mates (presumably females) would automatically prefer informative vocalizers over non-informative ones. This suggests that such mate choice would automatically drive more complex and ever more informative communication systems. However, a vast literature on sexual selection and sexually-selected communication systems is incompatible with this assumption. In fact, both theoretical considerations and a mass of empirical data suggest that the very nature of sexual selection drives sexually-selected signals inevitably toward exaggeration, false bluster, and outright deceit. Furthermore, sexual selection alone is clearly inadequate to explain the evolution of language: it is falsified by two everyday facts. First, sexually-selected traits are typically much more pronounced in the displaying sex. In mammals, virtually universally, this means males (hence it is males that sport antlers, manes, and bright colors, males that perform elaborate displays, and males that do the singing among most birds, whales, and seals). However, human language abilities are sexually egalitarian and, if anything, more highly developed in females. Second, sexually-selected traits are almost always expressed only at puberty, but language is remarkable for its precocity: a human ten-month-old already has remarkable language skills. Thus, from a

comparative perspective, it is clear that these facts are incompatible with the notion that sexual selection could drive meaningful communication.

I have proposed an alternative that, although it seems intuitive, appears to have remained unexplored by previous theorists. The idea is simple: that propositional meaning evolved in a context of communication among kin and particularly between parents and their offspring. I call this idea the "mother tongues" hypothesis. Primates in general and humans and apes in particular have an extremely long childhood, during which the offspring are completely dependent on their parents. Because of this, each individual offspring is of great importance in species of our hominid line. Indeed, chimps appear to be barely able to maintain current population levels, given their slow birthrate, and our early ancestors presumably faced a similar situation. Thus, a unique feature of the hominid line that differentiates us from other animals with complex vocalizations like birds or seals is the high premium on the individual offspring combined with a very long period of childhood learning, when a huge amount of information can be passed from parents to offspring. The animal world offers us ample testament that kin selection can drive parent/offspring communication to become honest, meaningful, and informative. Indeed, one of the few examples of propositional meaning in animal communication is alarm calls: vocalizations made to alert conspecifics, typically close kin, of the proximity of a predator. These calls seem to have evolved in many different species including birds, primates, and rodents largely in a context of kin selection. Perhaps the most striking system of kin communication is the dance language of honeybees, which allows one honeybee to convey highly accurate information about the location and quality of food, shelter, or water to another honeybee, invariably her sister (since all the bees in a hive share the same mother). The mother tongues hypothesis builds upon this comparative data, and unlike the sexual selection hypothesis, it is compatible with current evolutionary theory. What's more, two of the peculiarities of language are obviously understandable within this framework: the young age at which language acquisition starts, and the slight advantage of women's linguistic abilities over men's.

In closing, I want to consider an aspect of language that many commentators consider most central and most unique: syntax. I use this term, as do most theoretical linguists, to single out the particular and very highly-developed human ability to combine low-level units into hierarchically-structured larger units (syllables into words, and words into sentences). Syntax entails the ability both to perceive and learn the complex structure inherent in other's utterances and to generate novel, structured utterances oneself. This ability is

critical if we are to be able to create structures of adequate complexity to convey complex thoughts. Note that syntax goes far beyond simple rules of concatenation, or word order. Joining utterances together in a sequence is a commonplace in animal communication, and some empirical evidence supports the idea that different orders are interpreted differently by listeners. However, human syntax generates hierarchical structures that can be mapped onto structured meanings, going far beyond simple word-order rules. Such hierarchical structures, typically diagrammed by linguists with tree diagrams, are crucial in all human languages. Often this structure is not denoted via word order: in German, for example, such structures are indicated mainly by grammatical inflection (the infamous "der" vs. "den" vs. "dem" distinctions), and word order can be varied for thematic emphasis. I now will turn to the question of how such structure-making abilities might have evolved.

Contemporary theorists of language evolution generally agree that it is impossible to explain language evolution via a single factor. Thus, most of the theories currently in the offing are what I call "dual stage" theories, which posit an intermediate stage of hominid evolution in which some sort of "protolanguage" evolved and was selected and perfected. Beyond this basic agreement, however, theories vary considerably. One popular theory, advanced by Étienne Bonnot de Condillac and lately championed by Michael Corballis, is that early language was gestural, rather than (or at least more than) vocal. However, this theory leaves unexplained how the transformation to a fully vocal form (speech) took place, or what might have driven it, and adds nothing to the search for the path to meaning or to syntax. A second theory, promulgated by Derek Bickerton (who coined the term "protolanguage") suggests that, quite intuitively, protolanguage involved meaningful words, perhaps with quite a large vocabulary, but with no syntax. This fits our prejudice of how cavemen spoke: "Og big" or "big Og" meaning the same thing. While popular, this theory is not particularly explanatory: it takes the origin of meaning for granted, while leaving the origin of syntax quite mysterious. (Bickerton suggests that some aspects of syntax may have been inherited from a kind of primitive economic scorekeeping: who gave what to whom?) The intuitive appeal of the "words-first" theory derives mainly from the fact that this is more or less what children do ontogenetically. However, this appeal may be superficial. First, it is unlikely that a child's one-word utterance (say, "up", used as a request to be picked up) in fact maps onto an adult's understanding of that word (as a preposition of place). Second, human babies are born with modern human brains, already designed by long evolution to learn modern language. There is no guarantee in this case that "ontogeny recapitulates phylogeny" and indeed good reason to doubt it. Thus, despite the intuitive

appeal of "words-first" theories (which form the general class of "synthetic" theories of language evolution), it seems worthwhile to consider some alternatives.

A broad class of alternative theories, termed "analytic" theories, put the intuitive cart before the horse, suggesting that complex structures *preceded* meaning in our evolution. The particular hypothesis that I find most plausible and most consistent with the comparative data was advanced briefly by Rousseau and later independently elaborated by Darwin. It is that protolanguage, rather than being a simpler version of modern language with less structure and simpler propositional capacity, was structurally quite complex but lacked meaning almost entirely. In short, protolanguage was more like music than like modern language. Quoting Darwin, "primeval man, or rather some early progenitor of man, probably first used his voice in producing true musical cadences, that is in singing". While I believe the limitation to singing is too strict (the production of rhythms, with hands or other objects, is equally important), I find an updated version of Darwin's hypothesis, which I call the "prosodic protolanguage" hypothesis, rather compelling. According to this model, both modern music and modern language are descended from an ancestral ability that had features in common to both: "prosodic protolanguage". These features include considerable structural complexity, a dependence on the auditory channel, the ability to convey gesture and emotion, and a thoroughly cultural basis (they are both learned by imitation and transmitted across many generations). The crucial missing feature in this prosodic protolanguage was, by hypothesis, the ability to convey propositionally-structured meanings. This hypothesis suggests that complex syntactic structure (paralleling that found in music) preceded the use of this structure in meaningful, propositional discourse. Thus, the first step in language evolution was to evolve a complex, songlike communication system, quite analogous to that of birds or whales. Then, the work of mapping these complex sonic structures onto meanings happened in a second stage, driven, as I have already argued, by kin selection.

There are several reasons that biologists should find this hypothesis appealing. First of all, we are positing a theoretical entity (prosodic protolanguage) that is similar in form to known, existing communication systems like birdsong. It is clear that such songlike systems, lacking in meaning, can evolve, because it has happened repeatedly in many independent lines during vertebrate evolution. Sexual selection provides an obvious way to ratchet up the complexity of such systems. Thus, studies of bird and whale song can provide valuable empirical data for understanding the evolution of complex sonic structures in human language and the mechanisms underlying such structural capabilities. Further-

more, there is no theoretical problem with supposing that the evolution of prosodic protolanguage was driven by sexual selection, as in bird- and whale-song. According to this theory, it was only when the selection regime switched from sexual selection to kin selection that children needed to have their language abilities come "on line" at an early age, instead of at puberty. However, note that female birds of many species can and do sing, both alone or in duets with males, and that in "singing" primates like gibbons, both males and females sing (in complex duets that serve joint territorial defense). Thus, even the initial stages of prosodic protolanguage need not have been male-dominated.

Thus, as for vocal imitation, it may be that the best place to look for hierarchical structural abilities is not in nonhuman primates, but in birds or whales. Indeed, with Marc Hauser, I have gathered data suggesting that one species of nonhuman primate (cotton-top tamarins) are quite limited in their ability to perceive hierarchical structures in sounds. While other primates (e.g. chimps) may be more capable (the research has yet to be done), the most likely place to look for structure-parsing abilities is in species that themselves produce hierarchically-structured vocalizations, like some songbirds. Note, however, that whatever form the hierarchical structure in birdsong takes, and this remains poorly understood, it will still be considerably simple than human syntax at the sentential level, because of its lack of meaning. Such phenomena as movement or anaphora, which form the central problems in syntax for theoretical linguists, require a mapping between sound and meaning that neither birdsong nor, by hypothesis, prosodic protolanguage has. Thus, the level of syntax I am positing here is akin to phonological syntax in modern languages.

A second advantage of the prosodic protolanguage hypothesis is that, rather than relying on a hypothetical, long-extinct protolanguage that cannot be studied experimentally, this hypothesis directs our research attention quite squarely to music. Music is a human universal, and a rather puzzling one for evolutionary biologists given its tenuous connection to anything obviously adaptive. However, if we consider music as a kind of "living fossil" of an earlier stage of our evolution, many of its curious aspects become understandable. Unlike language, musical ability varies immensely among individuals, and although no human culture lacks music, many individuals lack even basic musical abilities. Despite this, most people enjoy music and indeed are willing to spend considerable time and money to experience it (as might be expected if music played a more crucial role in our recent evolutionary history). Comparative study of musical traditions (ethnomusicology) will teach us about the various contexts in which music is used (e.g. sexual selection vs. group synchronization or solidarity) and can provide important data toward an eventual "universal

grammar" of music. Methodologically, the variability of musical skill makes it perfect for empirical research on the genetic, neural, and behavioral correlates of musical ability: it is quite easy to find normal humans who either excel at or lack musical ability. This is in sharp contrast with language, where a "normal" human lacking language is virtually an oxymoron. It is important to stress that modern music is NOT identical to prosodic protolanguage, any more than modern chimps are identical to our shared ancestor with chimps (they have been separately evolving for six million years, just like us). With this caveat in mind, however, the study of our human "music faculty" can be expected to teach us much about the language faculty. Even if the prosodic protolanguage hypothesis is eventually falsified (if for instance, music and language are found to rely on fundamentally different computational mechanisms, implemented via disjunct neural mechanisms and controlled by independent genes), we will have learned much of value by exploring it.

To summarize my phylogenetic argument, I have argued that we can explain three critical components of language evolution – vocal imitation, syntax, and meaning – via a dual-stage hypothesis. First, vocal imitation and a basic combinatoric syntax (with hierarchical structure but no meaning) evolved during an initial protolanguage stage to create a relatively complex prosodic protolanguage that was shared in a community, passed on and elaborated over many generations, and probably played a role in mate choice, group solidarity, and perhaps synchronization of group activities. Although language does not fossilize and it is hard to say much about the precise timing of this stage, it is tempting to speculate that such a prosodic protolanguage characterized *Homo erectus*, the first hominid to expand out of Africa and to occupy most of the Old World. To the extent that sexual selection played the crucial role during this stage, comparative data suggest that males would have dominated this domain and that productive skills would be expressed around puberty. At the end of this evolutionary period, the stage was set for the evolution of true language: learned sounds of adequate complexity existed, along with complex thoughts to be expressed, but the connection between the two had not been made. In the second stage of language evolution, we entered a completely different "mother tongues" selective regime. Here, the need to communicate propositional information to children, thus increasing their chance of survival, predominated. This led to the precocial, female-biased distribution of contemporary linguistic skills. While the protolinguistic stage in this theory is closely analogous to birdsong, the evolutionary move to the second stage was driven by the low birthrate and peculiarly long period of childhood dependence typical of the human/ape line (it is hard to imagine that the month-long nestling period of most birds provides

adequate scope for detailed tutelage: from an evolutionary viewpoint, it is easier to lay a few more eggs or gather a bit more food).

To end, I want to return to the Paris Linguistic Society's ban. I hope to have convinced you in the course of this talk that there are many sources of solid empirical data that are directly relevant to language evolution. Foremost among them are comparative data, gathered from the many other living species that share our planet and are related to us at varying degrees. Thus, I think the "data-poor" situation that led to the century-old ban on the discussion of language evolution is over. Future workers in this field need to overcome the old assumption that discussions of language evolution provide an automatic license to speculate freely and to entertain whatever hypotheses might seem intuitively appealing. On the contrary, the contemporary database is already rich enough to exclude many currently popular theories, and we stand to learn much more if researchers in this field focus on gathering relevant comparative data. Convergent evidence from ethology, linguistics, neuroscience, and bioacoustics, along with the constraints imposed by modern evolutionary theory, will allow us to build theories that are consistent with existing data and point the way to gather new data that can test (and potentially refute) these theories. Although many of the ideas I have advanced above may turn out to be incorrect, they are at least consistent, and (more importantly) they highlight relevant areas for future empirical research. Thus, I hope you will agree that language evolution is not only a fascinating topic, but a topic ripe for empirical scientific exploration. I believe the next ten years of research on this topic, central to what it means to be human, could be quite exciting.