The Evolution of Speech and Language*

Philip Lieberman+
Haskins Laboratories, New Haven, Conn.

Although the view that human language is "unique" and is disjoint from the communications systems of all other animals is still current (Lenneberg, 1967), the research of the past century has demonstrated that Charles Darwin's theory of evolution through natural selection is essentially correct. Human language can be no more disjoint from the communications systems of other living animals than human respiration or human locomotion. The apparent uniqueness of human language, like the apparent uniqueness of fully bipedal locomotion, merely reflects the fact that the intermediate forms are extinct.

Human locomotion and human language both can be viewed as the result of gradual processes that evolved from phylogenetically simpler hominid ancestors. A human characteristic like bipedal locomotion structures virtually all aspects of human behavior. Tool use and tool manufacture, for example, are possible in Homo sapiens because our hands are free. Tool use and tool manufacture, of course, crucially involve the presence of cognitive factors. Without the human brain, bipedal locomotion would not be that useful. The evolution of both bipedal locomotion and the human brain mutually reinforced the evolution of the behavioral patterns of tool use and tool manufacture, which, in turn, placed greater selective advantages on both bipedal locomotion and enhanced cognitive abilities. It thus is both necessary and meaningful to discuss the evolution of human characteristics like bipedal locomotion and language in terms of the different factors that may have structured the selective factors resulting in the retention of the mutations that ultimately created Homo sapiens. These factors also are, and have been, operant in the evolution of other species. We thus can form and test hypotheses concerning the nature of human language and speech using data derived from other species.

I will discuss some of the factors that may be involved in the evolution of human language. These factors are necessarily linked; the presence of one particular factor is not, in itself, an explanation of the evolution of language. The absence, or lack of development, of one factor or another for modern Homo


+Also University of Connecticut, Storrs. After 1 July 1974 at Brown University, Providence, R. I.

sapiens would imply the presence of intermediate grades of language relative to the language of present-day humans. As Darwin (1859) pointed out, evolution proceeds in small steps. I will propose a model that involves the evolution of a number of interrelated factors that gradually derive hominid linguistic ability. I will necessarily have to limit the discussion of each factor, and the list of factors obviously will not be complete, but I will discuss some of the data that make each factor part of a scientific theory, a theory that can be tested and extended.

Factor 1. Speech and Language

Since the focus of this conference is the relationship between speech and language, I will start with these factors, though I do not intend to claim that language is impossible without speech. Human language appears to involve closely the constraints of human speech. However, as I will try to show in the discussion of some of the other factors, other forms of language are possible without the presence of the particular characteristics of human speech.

The special link between human speech and human language was recognized in the pioneering 19th century studies of Broca (1861) and Wernicke (1874). Broca found that lesions in a small area of the brain situated near the motor cortex in the left, dominant hemisphere of the brain impaired speech production and writing. The victims of the "aphasia" could still move their tongues, lips, etc. In some instances they could sing, but they had difficulty when they either spoke or wrote. Lesions in the area of the brain that has come to be known as Broca's area essentially interfere with the organization of the articulatory maneuvers that produce speech, the "programs," as well as the written symbols that represent speech. Wernicke in 1874 described and localized the complementary aspect of aphasia. He located an area of the brain near the auditory centers of the left, dominant hemisphere. Lesions in this area produced an aphasia in which the victim left out words, used the wrong syntax, or "lost" the proper phonetic spellings of words. The victims of lesions in Wernicke's area essentially lose part of the "dictionary" and the grammar that every human carries about in his (or her) head. Both of these areas of the brain can be regarded as evolved additions to parts of the brain that deal with the production of sound (for Broca's area) and the perception of sounds (for Wernicke's area). Lesions in Wernicke's area clearly involve much more than the mere perception of sound, just as lesions in Broca's area involve much more than the ability simply to move the tongue, lips, jaw, etc. The total linguistic ability of the victim is impaired. The siting of these areas near the parts of the brain that are directly concerned with auditory signals suggests that special neural mechanisms evolved matched to, and as a consequence of, vocal communication.

We can test this hypothesis with data derived from the study of other species. In recent years a number of electrophysiological and behavioral studies have demonstrated that various animals have auditory detectors that are "tuned" to signals of interest to the animal. Even "simple" animals like crickets appear to have neural units that code information about the rhythmic elements of their mating songs. The calling songs of male crickets consist of stereotyped rhythmic pulse intervals and females respond to conspecific males by their songs (Hoy and Paul, 1973).

Similar results have been obtained in squirrel monkey (Saimiri sciureus). Wollberg and Newman (1972) recorded the electrical activity of single cells in
the auditory cortex of awake monkeys during the presentation of recorded monkey vocalizations and other acoustic signals. Eleven calls, representing the major classes of this species' vocal repertoire, were presented along with tone bursts, clicks, and a variety of acoustic signals designed to explore the total auditory range of these animals. Extracellular unit discharges were recorded from 213 neurons in the superior temporal gyrus of the monkeys. More than 80 percent of the neurons responded to the tape-recorded vocalizations. Some cells responded to many of the calls that had complex acoustic properties. Other cells, however, responded to only a few calls. One cell responded with a high probability only to one specific signal, the "isolation peep" call of the monkey.

The experimental techniques necessary in these electrophysiological studies demand great care and great patience. Microelectrodes that can isolate the electrical signal from a single neuron must be prepared and accurately positioned. Most importantly, the experimenters must present the animals with a set of acoustic signals that explores the range of sounds that the animal would encounter in its natural state. Demonstrating the presence of neural mechanisms matched to the constraints of the sound-producing systems of particular animals is therefore a difficult undertaking. The sound-producing possibilities and behavioral responses of most "higher" animals make comprehensive statements on the relationship between perception and production difficult. We can only explore part of the total system of signaling and behavior. "Simpler" animals, however, are useful in this respect since we can see the whole pattern of the animal's behavior.

The behavioral experiments of Capranica (1965) and the electrophysiological experiments of Frishkopf and Goldstein (1963), for example, demonstrate that the auditory system of the bullfrog (Rana catesbeiana) has single units that are matched to the formant frequencies of the species-specific mating call. Bullfrogs are members of the class of Amphibia. Frogs and toads compose the order of Anura. They are the simplest living animals that produce sound by means of a laryngeal source and a supralaryngeal vocal tract. The supralaryngeal vocal tract consists of a mouth, a pharynx, and a vocal sac that opens into the floor of the mouth in the male. Vocalizations are produced in the same manner as in primates. The vocal folds of the larynx open and close rapidly, emitting puffs of air into the supralaryngeal vocal tract, which acts as an acoustic filter. Frogs can make a number of different calls (Bogert, 1960). These calls include mating calls, release calls, territorial calls which serve as warnings to intruding frogs, rain calls, distress calls, and warning calls. The different calls have distinct acoustic properties.

The mating call of the bullfrog consists of a series of croaks varying in duration from 0.6 to 1.5 sec. The interval between each croak varies from 0.5 to 1.0 sec. The fundamental frequency of the bullfrog croak is about 100 Hz. The formant frequencies of the croak are about 200 and 1400 Hz. Capranica (1965) generated synthetic frog croaks by means of a PVO speech synthesizer (Stevens, Bastide, and Smith, 1955). This is a fixed speech synthesizer designed to produce human vowels. It serves equally well for the synthesis of bullfrog croaks. In a behavioral experiment Capranica showed that bullfrogs responded to synthesized croaks so long as the croaks had energy concentrations at either or both of these frequencies (200 and 1400 Hz). The presence of acoustic energy at other frequencies inhibited the bullfrogs' responses (joining in a croak chorus).

Frishkopf and Goldstein (1963) in their electrophysiologic study of the bullfrog's auditory system found two types of auditory units. They found cells
in units in the eighth cranial nerve of the anesthetized bullfrog that had maximum sensitivity to frequencies between 1000 and 2000 Hz. They found other units that had maximum sensitivity to frequencies between 200 and 700 Hz. The units that responded to the lower frequency range, however, were inhibited by appropriate acoustic signals. Maximum response occurred when the two units responded to time-locked pulse trains, at rates of 50 and 100 pulses per sec, that had energy concentrations at or near the formant frequencies of bullfrog mating calls. Adding acoustic energy between the two formant frequencies at 500 Hz inhibited the responses of the low-frequency single units.

The electrophysiologic, behavioral, and acoustic data are complementary. Bullfrogs have auditory mechanisms structured to respond specifically to the bullfrog mating call. Bullfrogs don't simply respond to any sort of acoustic signal as though it were a mating call. They respond only to particular calls that can be made only by male bullfrogs, and they have neural mechanisms structured in terms of the species-specific constraints of the bullfrog sound-producing mechanism. Capranica tested his bullfrogs with the mating calls of 34 other species of frogs. The bullfrogs responded only to bullfrog calls; they ignored all other mating calls. The croaks must have energy concentrations equivalent to those produced by both formant frequencies of the bullfrogs' supralaryngeal vocal tract. The stimuli furthermore must have the appropriate fundamental frequency.

The bullfrog has one of the simplest forms of sound-making systems that can be characterized by the Source-Filter Theory of sound production (Fant, 1960; to be discussed more fully below). His perceptual apparatus is demonstrably structured in terms of the constraints of his sound-producing apparatus and of the acoustic parameters of the Source-Filter Theory, the fundamental frequency and formant frequencies.

Factor 2. Plasticity and the Evolution of Human Speech

Frogs are rather simple animals but they nonetheless have evolved different species-specific calls. Some of the 34 species whose mating calls failed to elicit responses from Rana catesbeiana were closely related. Others were more distantly related. Clearly, natural selection has produced changes in the mating calls of Anuran species. The neural mechanisms for the perception of frog calls are at the periphery of the auditory system. They apparently are not very plastic since Capranica was not able to modify the bullfrogs' responses over the course of an 18-month interval. Despite this lack of plasticity, frogs have evolved different calls in the course of their evolutionary development.

Primates appear to have more flexible and plastic neural mechanisms for the perception of their vocalizations. Recent electrophysiologic data (Miller, Stutton, Pfingst, Ryan, and Beaton, 1972) show that primates like rhesus monkey (Macaca mulata) will develop neural detectors that identify signals important to the animal. Receptors in the auditory cortex responsive to a 200 Hz sine wave were discovered after the animals were trained by the classic methods of conditioning to respond behaviorally to this acoustic signal. These neural detectors could not be found in the auditory cortex of untrained animals. The auditory system of these primates thus appears to be plastic. Receptive neural devices can be formed to respond to acoustic signals that the animal finds useful.
Factor 3. Special Supralaryngeal Vocal Tract Anatomy

Modern man's speech-producing apparatus is quite different from the comparable systems of living nonhuman primates (Lieberman, 1968; Lieberman, Klatt, and Wilson, 1969; Lieberman, Crelin, and Klatt, 1972). Nonhuman primates have supralaryngeal vocal tracts in which the larynx exits directly into the oral cavity (Negus, 1949). In the adult human the larynx exits into the pharynx. The only function for which the adult human supralaryngeal vocal tract appears to be better adapted is speech production. Understanding the anatomical basis of human speech requires that we briefly review the Source-Filter Theory of speech production (Fant, 1960). Human speech is the result of a source, or sources, of acoustic energy being filtered by the supralaryngeal vocal tract. For voiced sounds, that is, sounds like the English vowels, the source of energy is the periodic sequence of puffs of air that pass through the larynx as the vocal cords (folds) rapidly open and shut. The rate at which the vocal cords open and close determines the fundamental frequency of phonation. Acoustic energy is present at the fundamental frequency and at higher harmonics. The fundamental frequency of phonation can vary from about 80 Hz for adult males to about 500 Hz for children and some adult females. Significant acoustic energy is present in the harmonics of fundamental frequency to at least 3000 Hz. The fundamental frequency of phonation is, within wide limits, under the control of the speaker who can produce controlled variations by changing either pulmonary air pressure or the tension of the laryngeal muscles (Lieberman, 1967). Linguistically significant information can be transmitted by means of these variations in fundamental frequency as, for example, in Chinese where these variations are used to differentiate among words.

The main source of phonetic differentiation in human language, however, arises from the dynamic properties of the supralaryngeal vocal tract acting as an acoustic filter. The length and shape of the supralaryngeal vocal tract determines the frequencies at which maximum energy will be transmitted from the laryngeal source to the air adjacent to the speaker's lips. These frequencies, at which maximum acoustic energy will be transmitted, are known as formant frequencies. A speaker can vary the formant frequencies by changing the length and shape of his supralaryngeal vocal tract. He can, for example, drastically alter the shape of the airway formed by the posterior margin of his tongue body in his pharynx. He can raise or lower the upper boundary of his tongue in his oral cavity. He can raise or lower his larynx and retract or extend his lips. He can open or close his nasal cavity to the rest of the supralaryngeal vocal tract by lowering or raising his velum. The speaker can, in short, continually vary the formant frequencies generated by his supralaryngeal vocal tract. The acoustic properties that, for example, differentiate the vowels [a] and [i] are determined solely by the shape and length differences the speaker's supralaryngeal vocal tract assumes in articulating these vowels. The situation is analogous to the musical properties of a pipe organ, where the length and type (open or closed end) of pipe determines the musical quality of each note. The damped resonances of the human supralaryngeal vocal tract are, in effect, the formant frequencies. The length and shape (more precisely the cross-sectional area as a function of distance from the laryngeal source) determine the formant frequencies.

The situation is similar for unvoiced sounds where the vocal cords do not open and close at a rapid rate, releasing quasiperiodic puffs of air. The source of acoustic energy in these instances is the turbulence generated by air rushing through a constriction in the vocal tract. The vocal tract still acts as an
acoustic filter but the acoustic source may not be at the level of the larynx as, for example, in the sound [s] where the source is the turbulence generated near the speaker's teeth.

The anatomy of the adult human supralaryngeal vocal tract permits modern man to generate supralaryngeal vocal-tract configurations that involve abrupt discontinuities at its midpoint. These particular vocal-tract shapes produce vowels like [a], [i], and [u], which have unique acoustic properties. The acoustic properties of these sounds minimize the problems of precise articulatory control. A speaker can produce about the same formant frequencies for an [i], for example, while he varies the position of the midpoint area function discontinuity by 1 or 2 cm (Stevens, 1972). They are also sounds that are maximally distinct acoustically. They, moreover, are sounds that a human listener can efficiently use to establish the size of the supralaryngeal vocal tract he is listening to. This last property relates to Factor 1, the specialized speech decoding that characterizes human language. The reconstructions of the supralaryngeal vocal tracts of various fossil hominids that Edmund S. Crelin has made (Lieberman and Crelin, 1971; Lieberman et al, 1972; Lieberman, in press) indicate that some extinct hominids lacked the anatomical basis for producing these sounds, while other hominids appear to have the requisite anatomical specializations for human speech.

Factor 4. Syntactic Encoding and Decoding

There are three interrelated aspects to the cognitive abilities that underlie language: syntactic encoding and decoding, automatization, and logical ability. Syntactic encoding and decoding obviously involves the presence of neural mechanisms. Although we don't know very much about the workings of the brain, we don't have to know how the brain works to know what it does. A transformational grammar (Chomsky, 1957, 1964, 1968) is, among other things, a formal description of the syntactic encoding that is a characteristic of human language. Encoding in a more general sense seems to be a characteristic of other forms of human behavior.

A grammar to a linguist is not a set of prescriptive rules for writing sentences. A grammar is instead a formal description of some aspect of linguistic behavior. As Chomsky (1957:11) puts it:

Syntactic investigation of a given language has as its goal the construction of a grammar that can be viewed as a device of some sort for producing the sentences of the language under analysis. More generally, linguists have been concerned with the problem of determining the fundamental underlying properties of successful grammars.

The fundamental property of grammar that Chomsky revealed is its "transformational syntax." Chomsky demonstrated that language must be viewed as a two-level process. Underlying the sequence of words that constitutes a normal, grammatical sentence is a "deep phrase marker" (Chomsky, 1964), which is closer to the logical level of analysis necessary for the semantic interpretation of a sentence. The transformational syntax is the "device" that restructures the deep, underlying

---

1 As well as consonants like [g] and [k] which involve the velar region of articulation.
level of language that is suited for semantic analysis, into the actual sentence that a person writes or speaks. The aspect of transformational syntax we want to stress is its encoding property, which is formally similar to the process of speech encoding (Liberman, 1970).

In Figure 1 we have presented a diagram that is essentially similar to the "parsing" or "constituent analysis" (Bloomfield, 1933) of traditional grammarians. The symbol S stands for sentence, NP for noun phrase, VP for verb phrase, V for verb, N for noun, and T for article. The diagram shows the syntactic relationships of the words of the sentence The man is old. The words the man, for example, constitute a noun phrase, the words is old constitute a verb phrase, which in turn is made up of a verb plus a second noun phrase. The word old constitutes the second noun phrase (the article of the second noun phrase reduces to an implied article). Diagrams of this sort are quite traditional. The first noun phrase could be called the subject of the sentence, the second, the object or predicate, etc. Semantic relationships are often "explained" by means of diagrams of this sort. The "actor-object" relationship, for example, is apparent in the diagram of the sentence Joe hit the man. The actor is the noun preceding the verb, the object the noun following the verb. We have simplified these diagrams and many of the details that a grammarian might find essential have been eliminated, but the essential facts and "explanatory" power of these diagrams have been preserved. Parsing is a "device" that formally "explains" some aspects of semantics, i.e., it reduces semantic analysis to a mechanical procedure. The noun to the left of the verb is the actor, that to the right of the verb is the object,
i.e., the noun acted on. The interesting thing about human language is that no one ever really utters sentences like *Joe hit the man* and *The man is old* when he wants to convey the information in the sentence *Joe hit the old man*. The two underlying deep phrase markers that would result in the simple sentences *The man is old* and *Joe hit the man* are encoded, i.e., scrambled together into one more complex sentence. The process is general and pervasive. The sentence *Joe hit the dirty old man who was wearing the red hat* would have underlying it a set of deep phrase markers that could have resulted in the sentences: *Joe hit the man. The man is old. The man is dirty. The man was wearing a hat. The hat is red.* It's much faster to utter the single complex sentence than the set of simpler sentences underlying it. The listener also doesn't have to keep track of the semantic referents and remember that we're talking about the same man in the first four simple sentences. All four repetitions of the word *man* are collapsed into a single *man* in the complex sentence. The two repetitions of the word *hat* are collapsed into a single *hat*. The complex sentence has fewer words and doesn't require keeping track of the semantic referents of the six "simple" sentences.

The transformational syntax can be regarded as the device that rearranges, deletes, and adds words to form the sentences of human language. The transformational syntax makes it impossible to sort out mechanically the semantic relationships of the words of complex sentences by using traditional sentence parsing. The "actor-acted on" relationship, for example, is semantically equivalent in the sentences *Joe hit Bill* and *Bill was hit by Joe*, though the words are on opposite sides of the verb. There are a number of reasons why traditional constituent grammars are not, in themselves, able to account for the properties of human languages (Chomsky, 1957, 1964; Postal, 1968), but it's enough to point out that they cannot account for the syntactic encoding that is characteristic of human language and for the complementary decoding that must take place when a listener or reader interprets a sentence.

**Factor 5. Automatization**

Human language involves rapidly executing complex sequences of articulatory movements or making equally complex perceptual decisions about the identity of particular sound segments. At a higher level, complex syntactic relationships must be determined. None of these processes is, however, what the speaker or listener is directly concerned with. The semantic content of the message is the primary concern of the speaker or listener. The sending and receiving processes are essentially automatic. No conscious thought is expended in the process of speech production, speech perception, or any of the syntactic stages that may intervene between the semantic content of the message and the acoustic signal. It is clear that "automatized" skills are not unique to human language. Other aspects of human activity, such as dance for example, involve similar phenomena. The novice dancer must learn the particular steps and movements characterizing a particular dance form. Once the steps have been learned they become automatized. The dance itself involves the complex sequences. Playing the violin, skiing, or driving a car all involve automatized behavior.

The bases for the automatized behavior that is a necessary condition for human language may reside in cross-modal transfers from other systems of hominid and primate behavior. Tool use, for example, requires a high degree of automatization. You can't stop to think how to use a hammer every time you drive a nail in. Hunting is perhaps a still stronger case. A successful hunter must be
able to thrust a spear or throw a stone without pausing to think about the mechanics of spear thrusting or stone throwing. Natural selection would quickly favor the retention of superior automatization. Automatized behavior pervades all aspects of culture. Indeed a cultural response is, to a degree, a special case of automatized behavior. Electrophysiological data derived from rhesus monkey demonstrates that automatization in primates involves establishing special pathways in the animal's motor cortex as the animal "learns" to perform a task (Evarts, 1973). Evarts observed the electrical activity of motor cortex neurons and the animal's muscles during the performance of learned hand movements. The animal's muscular activity when he learned to perform the task was extremely rapid. Its muscles acted within 30 to 40 msec, about twice as fast as it could have responded if it had to "think about" the task. Short response times like this usually are associated with reflex actions, but these short response times were the result of the animal's automatizing a response. The learned, automatized responses of simpler animals generally are not taken as tokens of the animal's "culture," but they nonetheless exist. The function of play in animals may indeed be to learn various patterns of automatized behavior germane to the animal's "culture." Puppies spend a lot of time staging mock battles, kittens stalk, etc. It wouldn't be difficult to devise appropriate experiments to explore the possible connection between play and automatization.

A special factor of automatized behavior may be that a "plastic" period is involved. It is comparatively easier to shape behavior during the plastic period. Afterwards, it is either impossible or more difficult for the animal or human to learn the automatized behavior. Puppies thus can be trained more readily than adult dogs. We're just beginning to appreciate some of the critical periods involved in learning various activities. Human newborns, for example, can be trained to walk alone about two months earlier than they normally do, if we take advantage of a critical period. Brief daily exercise of the walking reflexes that exist in human newborn leads to an earlier onset of walking alone (Zelazo, Zelazo, and Kolb, 1972). If a newborn infant is held under his arms and his bare feet are permitted to touch a flat surface, he will perform well-coordinated walking movements similar to those of an adult. This reflex normally disappears after about eight weeks. However, if the infant is actively exercised throughout this period, the reflex can be transferred intact from a reflex to a volitional action. Latent periods are quite important in the acquisition of human language (Lenneberg, 1967). All humans can readily learn different languages in their youth. They all appear to retain this ability to at least age 12 (Sachs, Lieberman, and Erickson, 1973). Most humans, however, can learn a foreign, i.e., unfamiliar, language only with great difficulty (or not at all) during adult life. There are, of course, exceptions to this rule and some adults are quite fortunate in retaining the ability to learn new languages with great facility. The same comments probably apply to learning to play the violin, tight-rope walking, etc., though no definitive studies have yet been made.

**Factor 6. Cognitive Ability**

Cognitive ability is a necessary factor in human language. Linguists often tend to assume that cognitive ability is linguistic ability. Indeed, since the time of Descartes the absence of human language in other animals has been cited as a "proof" of man's special status and of the lack of cognitive ability in all other species. Human language has been assumed to be a necessary condition for human thought. Conversely, the absence of human language has been assumed to be evidence of the lack of all cognitive ability.
The cognitive abilities traditionally associated with presumably "unique" human behavioral patterns like tool use and toolmaking have been observed in a number of different animals. Chimpanzees have often been observed using and making tools (Lawick-Goodall, 1972), but they are not the only primates who have been observed in the act of using and making tools. Beck (in press) reviews much of the evidence that shows tool use in other primates in their natural settings. Tool use has also been carefully documented in the sea otter (Kenyon, 1969). Sea otters float on their backs and use stones as anvils against which they break the shells of crustaceans. The sea otters will hold onto stones that are suitable anvils, tucking the stone under a flipper as they swim between meals. The sea otter thus not only uses a stone tool, but preserves it for future anticipated applications.

Tool use and toolmaking under less natural conditions has even been observed in birds. Laboratory-raised northern blue jays (Cyanocitta cristata) have been observed tearing pieces from pages of newspapers and using them as tools to rake in food pellets which were otherwise out of reach (Jones and Kamil, 1973). The toolmaking techniques that can be observed in living nonhuman animals are rather simple. The stone tools associated with the earliest known fossil hominids are, however, also rather simple. We'll discuss the cognitive implications of different toolmaking techniques, but it is clear that the tool-using and toolmaking behavior of many living animals is a reasonable approximation to the initial base on which natural selection acted in the gradual evolution of hominid behavior.

The linguistic ability of present-day chimpanzees also is evidence of the cognitive "base" that is present in living nonhuman animals. Chimpanzees do not have the speech-producing anatomy of modern Homo sapiens (Lieberman et al., 1972). They could not produce human speech even if they had the neural devices, localized in Broca's area, that organize the complex articulatory gestures of human speech. Chimpanzees, however, can be taught to use a modified version of American Sign Language. American Sign Language is not a method of "finger spelling" English words. It is instead a system that makes use of gestures that correspond to complete words, morphemes (e.g., past tense), or phrases (Stokoe, 1960). It has a different grammar than standard English and really is a different language with its own linguistic history. Chimpanzees taught this sign language communicate in a linguistic mode with human interlocutors (Gardner and Gardner, 1969; Fouts, 1973). They also can be observed communicating with other chimpanzees through sign language (Fouts, 1973). Other experimenters have taught chimpanzees to communicate with humans by means of plastic symbols (Premack, 1972) and by means of a computer keyboard (Rumbaugh, 1973). These experiments and observations demonstrate that chimpanzees can communicate in a linguistic mode. Chimpanzees, for example, are aware of what constitutes a "grammatical" syntactic construction (Rumbaugh, 1973). They conjoin words to form sentences such as I want apples and bananas, and they understand the principle of negation (Premack, 1972). They generalize the use of words, categorize in terms of semantic attributes, and use syntactic and logical constructs such as conditional sentences, Lucy read book if Roger tickle Lucy (Fouts, 1973). The chimpanzee's cognitive linguistic abilities are, at worst, restricted to some subset of the cognitive abilities available to humans. Chimpanzees may lack the syntactic encoding that must be formally described by a transformational syntax in human language. Definitive experiments investigating the syntax of chimpanzee communications using sign language have yet to be done, and we don't really know whether
their sentences are syntactically encoded. The difference at the cognitive level may, however, be quantitative rather than qualitative.

It is important to note, at this point, that quantitative functional abilities can be the bases of behavioral patterns that are qualitatively different. I think that this fact is sometimes not appreciated in discussions of gradual versus abrupt change. A modern electronic desk calculator and a large general-purpose digital computer, for example, may be constructed using similar electronic logical devices and similar magnetic memories. The large general-purpose machine will, however, have 1,000 to 10,000,000 times as many logical and memory devices. The structural differences between the desk calculator and general-purpose machine may thus simply be quantitative rather than qualitative. The "behavioral" consequence of this quantitative difference can, however, be qualitative. The types of problems that one can solve on the general-purpose machine will differ in kind, as well as in size, from those suited to the desk calculator. The inherent cognitive abilities of humans and chimpanzees thus could be quantitative and still have qualitative behavioral consequences.

An Interactive Model for the Evolution of Human Language

I have discussed some of the factors that I think are relevant to the evolution of language and speech. The first hominid "languages" probably evolved from communication systems that resembled those of present-day apes. The social interactions of chimpanzees are marked by exchanges of facial and body gestures as well as vocalizations (Goodall, 1968). Chimpanzees also use tools, make tools, and engage in cooperative behavior (for example, hunting). All of these activities have been identified as factors that may have placed a selective advantage on the evolution of enhanced linguistic ability (Washburn, 1968; Hill, 1972).

Australopithecus africanus (Lieberman, 1973, in press) essentially has the same supralaryngeal vocal tract as present-day apes. This, however, still would allow A. africanus to establish a vocal language if other prerequisites were also present. A. africanus would have had to have had the motor skills and automation necessary to produce the coordinated articulatory maneuvers that are necessary for the production of speech. Australopithecines were more advanced in relative brain size than any present-day ape, and, if external brain morphology means anything, they were more advanced in internal organization too. Quantities of shaped stones associated with early hominids have been recovered. These stones probably were used, among other things, as projectiles (Leakey, 1971). The transference of patterns of "automatized" behavior from activities like toolmaking and hunting would have facilitated the acquisition of the motor skill necessary to make these sounds. Enhanced communicative ability would, in turn, facilitate the use of tools. The process would be circular, a positive feedback loop in which each step enhances the adaptive value of the next step. Particular neural capacities may initially not have been innately present. That is, they may not have been in place at birth like the auditory detectors of frogs, which don't appear to involve much, if any, learning. The plasticity of the Australopithecine auditory system, however, surely would have been at least as great as that of present-day rhesus monkeys, dogs, chaffinches, etc.

The initial language of the Australopithecines thus may have had a phonetic level that relied on both gestural and vocal components. The system may have become more elaborate as factors like tool use, toolmaking, and social interaction became more important. The ability to control rage and sex is one of the factors
that makes human society possible (Hamburg, 1963). Language is probably one of the most important factors in reducing the level of aggressive behavior in human society. Social control is as important a factor as hunting in the evolution of human society (Washburn, 1969). The level of interaction between mother and child which can be noted in the vocal and gestural communications of chimpanzee, in which the mother is the primary agency of socialization (Lawick-Goodall, 1972), is a good example of this source for the increased selective advantage of communication. As hominid evolution diversified and larger-brained hominids appeared in the Homo habilis/erectus lineage, the selective advantages of linguistic ability would have increased.

The final crucial stage in the evolution of human language would appear to be the development of the bent, two-tube supralaryngeal vocal tract of modern man. Figure 3 shows a divergence in the paths of evolution. Some hominids like the classic Neanderthal fossils appear to have retained the communication system that was typical of the Australopithecines, perhaps elaborating the system, but retaining a mixed phonetic level that relied on both gestural and vocal components (Lieberman and Crelin, 1971). Other hominids appear to have followed an evolutionary path resulting in almost complete dependence on the vocal component for language, relegating the gestural component to a secondary, "paralinguistic" function. The process would have been gradual, following from the prior existence of vocal signals in the linguistic communication of earlier hominids.

The bent supralaryngeal vocal tract that appears in forms like present-day Homo sapiens and the Es-Skhul V fossil allows its possessors to generate acoustic signals that (1) have very distinct acoustic properties and (2) are easy to produce, being acoustically stable. These signals are in a sense optimal acoustic signals (Lieberman, 1970, 1973, in press). If vocal communications were already part of the linguistic system of early hominids, the mutations that extended either the range or efficiency of the signaling process would have been retained. At some later stage (that is, later with respect to the initial appearance of the bent, two-tube supralaryngeal vocal tract) the neural mechanisms necessary for the process of speech encoding would have evolved. The human-like supralaryngeal vocal tract would have initially been retained for the acoustically distinct and articulatorily stable signals that it could generate. The acoustic properties of the vowels [i] and [u] and the glides [y] and [w], which allow a listener to determine the size of a speaker's supralaryngeal vocal tract, would have preadapted the communication system for speech encoding.

The process of speech "decoding" appears to involve crucially the left hemisphere of the brain. When isolated vowels are, for example, presented dichotically to a human listener there is no right-ear advantage so long as the listener is responding to vowel stimuli that could have been produced by a single, unique vocal tract. If the vowel stimuli are instead derived from a set of different vocal tracts, a strong right-ear advantage is evident (Darwin, 1971). The listener has to make use of a perceptual recognition routine that normalizes the incoming signals in terms of the supralaryngeal vocal tracts that could have produced the particular stimuli. The neural modeling of this recognition routine apparently involves the left, dominant hemisphere of the listener's brain. The traditional mapping of areas like Broca's and Wernicke's areas in the left hemisphere of the brain reflects the result of a coherent evolutionary process in which the human brain evolved special, unique mechanisms structured in terms of the matched requirements of speech production and speech perception.
Figure 3: Tentative evolution of recent hominids with respect to human species.
The Uniqueness of Encoding

Although the speech of modern Homo sapiens is a fully encoded system, we can't assert dogmatically that other animals and, in particular, various fossil hominids, had completely unencoded systems of vocal communication. The acoustic basis of speech encoding rests in the fact that the pattern of formant frequency variation of the supralaryngeal vocal tract must inherently involve transitions. The shape of the supralaryngeal vocal tract cannot change instantaneously. If a speaker utters a syllable that starts with the consonant [b] and ends with the vowel [æ] his vocal tract must first produce the shape necessary for [b] and then gradually move towards the [æ] shape. Formant transitions thus have to occur in the [æ] segment that reflect the initial [b] configuration. The transitions would be quite different if the initial consonant were a [d]. The non-human supralaryngeal vocal tract can, in fact, produce consonants like [b] and [d]. Simple encoding could be established using only bilabial and dental consonant contrasts. The formant transitions would all be either rising in frequency in the case of [bae] or falling in frequency for [dae]. It probably would be quite difficult, if not impossible, to sort the various intermediate vowel contrasts that are possible with the nonhuman vocal tract, but a simple encoding system could be built up using rising and falling formant transitions imposed on a general, unspecified vowel [V]. The resulting language would have only one vowel [a claim that has often been made for the supposed ancestral language of Homo sapiens (Kuipers, 1960)]. The process of speech encoding and decoding and the elaboration of the vowel repertoire could build on vocal-tract normalization schemes that made use of sounds like [s], which also can provide a listener, or a digital computer program, with information about the size of the speaker's vocal tract. Vocal-tract normalizing information could also be derived perhaps by listening to a fairly long stretch of speech and then computing the average formant frequency range. The process would be slower than simply hearing a token of [i] or [u], but it would be possible.

There might have been a gradual path towards more and more encoding for all hominid populations as social structure and technology became more complex. The preadaptation of the bent, two-tube supralaryngeal vocal tract in some hominid populations would have, if this were true, provided an enormous selective advantage. In other words, there may not have been any single path towards the evolution of encoded speech. Fossil hominids like Neanderthal man may have had cognitive abilities equal to those of hominids like Es-Skhul V. However, the absence of a preadapted, bent, two-tube vocal tract would have prevented them from generalizing the encoding principle.

Tool Use, Grammar, and Encoding

As we noted earlier, linguists often tend to view human language as though it were disjoint from all other aspects of human behavior. A linguistic grammar is essentially a formal description, or rather a formal abstraction, of certain aspects of language. Linguists, in general, would not think of applying the formal apparatus of a linguistic grammar to some other kind of human behavior. However, it is apparent that other aspects of human, and indeed of nonhuman, behavior can be described using the same formal apparatus. Reynolds, for example, who studied the play activity of young rhesus monkeys (Reynolds, 1972), found that rhesus monkeys have a number of stylized basic gestural patterns. These patterns are all quite short. They each consist of a particular body posture and facial expression. Some of the basic patterns involve movements and vocalizations. The
basic play patterns are essentially "atomic" units that combine in certain regular ways to form play sequences. Some of the basic patterns occur only at the start of play sequences; they are "initiators." Others can occur only at the end of a play sequence; they are "terminators." Still other basic patterns occur within play sequences. The monkeys will break off play whenever a basic pattern occurs in the wrong position. There are, if we borrow the terminology of linguistics, "grammatical" play sequences and we can describe these sequences by means of "grammatical" rules.

The "grammatical" rules that appear to be appropriate for the description of rhesus monkey play sequences are those usually associated with constituent analysis, i.e., sentence parsing. Let's consider the following short example of sentence parsing drawn from Chomsky (1957). Consider the following set of grammatical "rules."

(1) \( S \rightarrow NP + VP \)
(2) \( NP \rightarrow T + N \)
(3) \( VP \rightarrow V + NP \)
(4) \( T \rightarrow \text{the} \)
(5a) \( N \rightarrow \text{man} \)  
(5b) \( N \rightarrow \text{house} \)  
(5c) \( N \rightarrow \text{ball} \)
(6a) \( V \rightarrow \text{hit} \)  
(6b) \( V \rightarrow \text{lost} \)

Each rule \( X \rightarrow Y \) is to be interpreted as the instruction "rewrite \( X \) as \( Y \)." We can call the sequence of operations that follows a "derivation" of the sentence, The man hit the ball, where the number at the right of each line in the derivation refers to the "rule" of the grammar used in constructing that line from the previous line.

\[
\begin{align*}
S & \rightarrow NP + VP & (1) \\
T + N + VP & \rightarrow NP + VP & (2) \\
T + N + V + NP & \rightarrow T + N + VP & (3) \\
\text{the} + N + V + NP & \rightarrow T + N + V + NP & (4) \\
\text{the} + \text{man} + V + NP & \rightarrow \text{the} + \text{man} + V + NP & (5a) \\
\text{the} + \text{man} + \text{hit} + NP & \rightarrow \text{the} + \text{man} + \text{hit} + NP & (6a) \\
\text{the} + \text{man} + \text{hit} + T + N & \rightarrow \text{the} + \text{man} + \text{hit} + T + N & (2) \\
\text{the} + \text{man} + \text{hit} + \text{the} + N & \rightarrow \text{the} + \text{man} + \text{hit} + \text{the} + N & (4) \\
\text{the} + \text{man} + \text{hit} + \text{the} + \text{ball} & \rightarrow \text{the} + \text{man} + \text{hit} + \text{the} + \text{ball} & (5c)
\end{align*}
\]

Thus, the second line of the derivation is formed from the first line by rewriting sentence as \( NP + VP \) in accordance with rule (1), the third line is formed from the second line by using rule (2), etc. We could represent the derivation by means of the following diagram.
If we add a filter condition to the rules of the grammar it will mechanically derive a number of "grammatical" English sentences, e.g., The man lost the house, The man hit the house, etc. The filter condition states that no derivation shall be considered complete unless all of the alphabetic symbols are replaced by English words. The application of a particular rule in this grammar is contingent on only one fact—the left-hand symbol of a rule must be present on the last line of the derivation.

The grammar that we have discussed is what linguists call a "phrase structure" grammar. It's the formal embodiment of traditional sentence parsing. Phrase structure grammars in themselves cannot capture the encoded nature of the syntax of human language. Phrase structure rules, however, do have a role as a component of the grammar of human language (Chomsky, 1957, 1964). They have one formal property that, though it superficially appears trivial, is an important limitation of their explanatory power. A phrase structure rule can be applied in a derivation whenever the alphabetic symbol on the left of the rule appears on the last line of the derivation. A phrase structure rule thus can apply to a line of a derivation without considering its past history.

After digressing on the play activity of rhesus monkey and on phrase structure rules, we can now return to the question of the language of Neanderthal hominids. In fact, we have not really been digressing since the point that we want to make is that we can apply the "rules" of grammar to the analysis of some of the artifacts of Neanderthal culture, the stone tools and toolmaking techniques.

Stone Toolmaking Techniques and Encoding

The Paleolithic, or Old Stone Age, encompasses a period of perhaps almost three million years. There are important differences in the types of stone tools found in different parts of this era. The first tools, which are associated with the Australopithecines and Homo habilis, are either unshaped stones or stones that have a flake or two taken off them. The tools become progressively more complex and their manufacture ultimately involved taking many, many chips out of the piece of stone that the toolmaker started with. We might think of a process in which toolmakers continued to refine the process of tool fabrication, making the chips smaller and more refined as time went on. The basic technique, however, would be unchanged though new modifications would be introduced. The process would simply become more refined.
The technique involved in making these tools is conceptually similar to the process of whittling on a stick. You start by making an initial chip, then a second, a third, etc. In making a particular chip you have to keep only two things in mind: (1) the last chip that you made, and (2) the final form of the tool that you're trying to make. The process formally reduces to the phrase structure grammar with a filter condition that we just discussed. The filter condition is formally equivalent to stating that you know what sort of tool you're aiming for. The phrase structure grammar formally embodies the fact that you only need to know the last "line of the derivation," i.e., the state of the tool blank at the instant that you chip it. You don't need to have a memory of the operations involved in getting to that stage.

We would be wrong in thinking that all stone tools involved the same technology. About 600,000 years ago a radically different stoneworking technology started. The Levallois flake tools (Bordes, 1968) are the result of a multistage process. The toolmaker first prepares a core (Figure 5), a process which involves a number of steps itself to produce the basic shape. Once the core is ready, the toolmaker switches his technique. He chips out complete flakes, each of which may serve as a completed tool, with every blow of his hammer. The Levallois toolmaking technique cannot be reasonably described by means of a phrase structure grammar. A transformational grammar which formally incorporates a memory is necessary. There is no simple invariant "last chip" at which the toolmaker abruptly stops preparing the core and switches to flaking off the final products. The toolmaker rather has to keep in mind a particular functional attribute of the striking platform which involves the entire upper surface of the core (Bordes, 1968:27, 28). The formal "grammatical" description of the process must also reflect this degree of abstraction, which cannot be keyed to the appearance of a single "alphabetic" symbol that represents a particular chip of stone.

Phrase structure grammars cannot formally account for the syntax of human language (Chomsky, 1957, 1964); they also cannot serve as grammars of the Levalloisian tool technique that is one of the characteristics of the culture of Neanderthal man. Transformational grammars, as we noted, introduce the concept of encoding into syntax. Although we cannot positively conclude that the grammar of the syntax of Neanderthal language had a transformational component, their Levalloisian stone tools suggest a degree of cognitive development that formally calls for a transformational grammar. Many other aspects of the culture of modern human populations need transformational descriptions if we attempt to derive a formal description. Marriage customs, for example, involve constraints on the lineages of both bride and groom that include a memory component. Death rituals involving funeral goods also implicitly require some knowledge of the former life and habits of the corpse.

The most likely assessment of the encoding abilities of Neanderthal man thus would be that language was encoded, but not nearly as encoded as modern Homo sapiens'. The development of the Neanderthal supralaryngeal vocal tract was not suitable for fully encoded speech. The neural structures of the brain that play so crucial a role in the perception of encoded speech in the dominant, left hemisphere of the brain would therefore probably not have been as well developed in Neanderthal man. Language, however, would exist though it would not be the language of modern Homo sapiens. Language, like other human attributes, appears to be the result of a gradual evolutionary process, and intermediate stages and
Figure 5: Paleolithic stone tools to illustrate an analogy between early tool-making and syntax in language. (a) Handaxe or coup de poing. Lower Paleolithic of Africa, Europe, southwestern and southern Asia. Length of average specimens approximately 10 to 15 cm. Similarly chipped on opposite surface. (b)-(e) Stages in the manufacture of a Levallois flake from a prepared tortoise core. Late Lower Paleolithic and Middle Paleolithic, in many of same geographic areas as (a) above. Approximately similar scale. (e) The prepared tortoise core, back and side view; the flattish underside is also chipped. [After G. W. Hewes (this conference) "Some Comments on Mattingly's Paper and on Levallois Flake Tools."]
common underlying factors are to be expected in the languages and communications systems of extinct earlier hominids and of other living species.

REFERENCES


Lawick-Goodall, J. (1972) In the Shadow of Man. (New York: Macmillan).


44