THE BEGINNINGS OF SPEECH

Michael Studdert-Kennedy*

INTRODUCTION

Man's life is diverse. The range of habitats, natural and man-made, to
which he has adapted is incomparably wider than that of any other species.
This is so because there evolved in man capacities for rapid cultural
evolution to augment the lengthy biological processes of adaptive radiation.
These capacities have permitted him to create new and unpredictable patterns
of behavior in the face of both old and new contingencies. The nature of
these capacities is quite unknown. But we can be sure that language is among
them, and that an understanding of its biology would take us a long way toward
understanding the history of man and of the earth during the past 10,000
years.

Unfortunately, "...the development of human speech represents a quantum
jump in evolution comparable to the assembly of the eucaryotic cell" (Wilson,
1975, p. 556). Whatever the lost links in phyletic evolution since the first
hominids diverged from the apes, presently living species offer few analogies
and even fewer homologies with language. In fact, the most fruitful
approaches to its biology seem to be those that have been followed for many
years by developmental psycholinguists (for reviews, see Brown, 1973; Dale,
1976; Ferguson & Slobin, 1973) and by students of neurophysiology (e.g.,
Lenneberg, 1967; Lenneberg & Lenneberg, 1975; Whitaker & Whitaker, 1976):
first, study of its ontogeny, with particular attention to similarities within
and across language communities; second, study of its pathology in childhood
and adult disorders.

The present chapter makes no attempt to review the vast, resulting
literature. Instead it undertakes to examine, critically, several tempting
analogies with language in the great apes and in the song-learning of certain
birds. Analogies often have the heuristic value of leading us to look at
familiar facts from a fresh viewpoint. Moreover, they may be instructive even
if they prove to be false.

*Also Queens College and Graduate Center, City University of New York.

Acknowledgment: Most of the work on this chapter was done while the author
was on sabbatical leave as a visiting scientist at the Center for Interdiscipli-
nary Research, University of Bielefeld, Germany, and a participant in a
nine-month conference of anthropologists, biologists, and psychologists on
Early Development in Animals and Man (September 1, 1977 through May 31,
1978). The chapter will be published in G. B. Barlow, K. Immelmann, M. Main,
and L. Petrinovich (Eds.), Behavioral Development: The Bielefeld

[HASKINS LABORATORIES: Status Report on Speech Research SR-58 (1979)]
THE NATURE OF LANGUAGE

If we compare language with other animal communication systems, we are struck by its breadth of function. The flashing white rump of the fallow deer denotes alarm; the "peep" of the squirrel monkey indicates that it is alone and wishes it wasn't; the "song" of the chaffinch informs the interested listener of its species, sex, local origin, personal identity and readiness to breed or fight. Even the elaborate "dance" of the honey bee merely conveys information about the direction, distance and quality of a nectar trove. But language can convey information about all these matters and many more besides. In fact, it is the peculiar property of language to set no limit on the possible topics of reference.

More exactly, no language consists of a finite number of sentences. This may be demonstrated by formal proof (Chomsky, 1956), or by the persuasive calculation that a single rendering of all grammatical English sentences of up to, say, twenty words in length would last longer than the history of the earth (Miller, Galanter, & Pribram, 1960, p. 146). In fact, no normal speaker of a language—no matter how limited his vocabulary or tedious his conversation—speaks by rote or constructs an utterance by drawing its components from a store of ready-made phrases.

How does language achieve this openness or productivity? There are several crucial features to its design (Hockett, 1960). First, language is learned: It develops under the control of an open, rather than of a closed genetic program (Mayr, 1974). Transmission of the code from one generation to the next is therefore discontinuous: Each individual recreates the system for himself. There is ample room here for creative error—probably a central factor in the evolution of language and in the constant process of change that all languages undergo (Kiparsky, 1968). One incidental consequence of this freedom is that the universal properties of language (whatever they may be) are largely masked by the surface variety of the several thousand languages now spoken in the world, not to mention their thousands of dialects and idiolects.

A second condition of productivity is that linguistic signals are arbitrary. With a few onomatopoetic exceptions, only by coincidence does a sign share any property with its referent. Of course, many other animal signals are arbitrary: the courtship rituals of the great-crested grebe, the red spot of the courting stickleback, the flush of a shamed human. But under the surface of such instances, some unknown physiological necessity is at work. These are not the arbitrary signs of convention by which bird, oiseau, Vogel and uccello are equivalent. Notice that if signs were iconic rather than arbitrary, the number of possible referents would be limited by the signaling organism's physical capacity to represent or depict.

A third, closely related condition of productivity is that signals are discrete rather than analog. To be precise, signals are perceived as discrete, even if they are not physically separable. Here again, if signals were not categorized by the receiver and if changes of meaning required changes of degree along some continuous scale, the number of possible signals...
would be limited by the number of possibly and perceptibly variable dimensions of the signal.

A final condition of productivity, and the one to which we will give most attention, is that language has two hierarchically related levels of structure: its signal elements are combined according to two more-or-less independent systems of rules. At the lower level of each language, the phonology or sound system, a small set (usually between 20 and 60) of meaningless phonemes (consonants and vowels) is specified, together with rules for their combination into morphemes (meaningful units which, for present purposes, we may treat as roughly equivalent to words). These are the rules that permit a vast, if not infinite, lexicon to be constructed by permutation and combination of a few dozen "alphabetic" units.

At a second level of structure, that of syntax, are specified the rules for combining words into meaningful sentences. These are the rules that permit us to predicate relations among objects or events. Central to the syntax of every known language are "recursive rules" by which a sentence may be treated as a component in another sentence. This capacity to embed a sentence within a sentence means that the set of all possible sentences in a language is infinite (Chomsky, 1956). Moreover, it is through this device that we can extend our communicative reach by constructing complex, sentential "names" for referents not represented in our lexicon, a trick already in the armory of many 3-year-olds: "I want the one Mary's got" (Limber, 1973). Incidentally, it is this central, inventive (though commonplace) use of language that Premack (1976, p. 15) thinks it "absurd" to expect of the chimpanzee.

IMPLICATIONS OF DUAL STRUCTURE

We begin to apprehend the importance of a dual structure, if we imagine a language with only one level, say that of sound (cf. Liberman & Studdert-Kennedy, 1978). Such a language would consist of meaningless elements (perhaps consonants and vowels) combined into lexical items, a set of "words" each with a different referent. Its users would presumably be confined to ostensive definition. For even if they were able to conceive of absent objects ("The bear we met yesterday") or abstract ideas ("The solar year") and were able to construct, from their phonetic resources, new lexical items to refer to them, they would be quite unable, lacking discursive speech, to establish the new meanings with their fellows. It is only by means of syntax that we are able to deploy old (known) words into new (previously unknown) statements—such as those that define new words. In short, rules for syntactic structure are a sine qua non of linguistic productivity.

The lack of a sound structure, on the other hand, would be less crippling. For, even if we were to replace every word in the lexicon with an arbitrary number (as might be done if the lexicon were stored in a computer), the syntactic structure of any particular utterance would be preserved despite the total loss of phonetic equivalences. (It is for this reason that linguists sometimes describe a language as an abstract system of communication, independent of its medium of expression.) Each lexical item would then be a totally distinct sign, lacking any systematic physical relation to any other. Of course, the number of such irreducible, holistically distinct
signals that humans are capable of recalling, producing and identifying at even a moderate rate—let alone the 50 bits/second typical of much speech—is certainly small, and it is not surprising that most vertebrate communication systems dispose of no more than 10 to 40 signals (Wilson, 1975, p. 183). However, a small lexicon does not preclude a productive syntax. That is why Premack (1976) and Rumbaugh (1977) saw no need for a formational structure in the visual symbols they devised for their pongid pupils.

Nonetheless, having granted that phonological (or word formational) structure is not, in principle, necessary for productive language, we must next acknowledge that every known language does, in fact, display it. The "extra" level of sound structure—which perhaps was prior to syntax in phyletic evolution, as it is in ontogeny—must therefore fulfill some function.

That function, as we have already suggested, is to facilitate the formation of a lexicon. Whether or not the lexical, or "naming," function is at the root of language, as is sometimes argued (e.g., Lancaster, 1968), most linguistic communities do have—in addition to their everyday lexicon of several thousand words—large, more-or-less specialized vocabularies, crucial to their cultural elaboration of the environment. This is as true of "primitive" peoples, such as the Hanunoo of the Philippines with their vast inventories of flora and fauna (Levi-Strauss, 1968) as of a modern industrial society with its proliferation of technical terms and subculture jargon. Thus, the seemingly trivial discovery that an essentially unlimited lexicon could be constructed from a small "alphabet" of sounds may have been the catalyst that set linguistic development in motion by providing an interface between man's intellect and his peripheral anatomic structure (Liberman, 1970; Mattingly, 1975). Certainly, it is at the level of the signaling system (that is, of speech) rather than of the abstract syntactic and semantic structure, that we find the clearest traces of biological adaptation, and it is therefore primarily with speech that the following sections are concerned.

THE SIGNALING SYSTEM

The sounds of any language can be viewed as the product of a sound source and a resonant filter. The sound source is usually either the "voice" produced by rapid pulsing of the vocal cords (as in the final sounds of "be" and "do"), the hiss of air blown through a narrow constriction (as in the initial and final sounds of "safe" and "thrush") or both (as in the final sounds of "leave" and "bees"). The resonant filter is the vocal tract, that is, the cavities of the pharynx, mouth, and nose.

The pulsing of the vocal cords at fundamental frequencies of roughly 90 to 250 Hz for males, 150 to 350 Hz for females and somewhat higher for small children, yields a signal rich in harmonic frequencies (multiples of the fundamental). Relatively slow variations in fundamental frequency over the course of an utterance yield the characteristic melody or intonation of speech. Taken with systematic variations in intensity, rate and rhythm, this melody is the basis of speech prosody, and plays an important role in communicating the emotional tone of an utterance, as well as, to some extent, its syntactic structure (e.g., question, statement, imperative). To the
unfamiliar listener (whether infant or foreigner) the slow variations of
prosody are probably more salient than the rapid patter of consonant-vowel
syllables. But it is primarily by syllables that the distinctively linguistic
(lexical and syntactic) information is carried. That, incidentally, is why
writing systems encode phonetic segments, but not prosody.

For the most part, this distinctively linguistic information is conveyed
by systematic variations in the "tuning" of the vocal tract. The curved
column of air in the tract, like that in an Alpine horn, resonates in
characteristic frequency bands (or formants) when set in motion by air from a
vibrating source, with the result that some of the source frequency components
are amplified, while others are attenuated. If we vary the size and shape of
the resonating tract by shifting the relative positions of the articulators,
especially the tongue, lips, jaw, and soft palate, the resulting shifts in the
formants yield the various sound spectra characteristic of particular phonetic
segments. The reader may find it instructive to monitor the position and
shape of his tongue as he runs it around the vowel triangle: eat, it, et, et,
aht, ought, oot.

THE SOUND PATTERN OF LANGUAGE

Here we must introduce the concept of a sound system or phonology. Each
language forms its words from a relatively small "alphabet" of distinctive
phonetic segments, termed phonemes. These are its consonants and vowels, and
in English there are about 35 of them, depending on dialect. The phonemes are
not chosen randomly. Each may be described in terms of the small set of
binary features (usually, a dozen or so) deployed in a particular language.
The phonemes may then be classified according to their shared features and the
resulting classes contrasted with one another on the basis of their feature
oppositions. A basic division, observed in every language, is between
consonants, formed by a more or less complete constriction of the vocal tract,
and vowels, formed with a relatively open tract. From their contrastive
combination is formed the fundamental unit of all spoken language, the
consonant-vowel syllable. It is the repeated opening and closing of the tract
and the consequent repetitive frequency and amplitude modulation, or syllabic
beat, that establishes the characteristic rhythms of human speech.

We may draw further contrasts among the phonemes (Figure 1). For
example, in English we may draw contrasts between voiced (/b,d,v,z/) and
voiceless (/p,t,f,s/), between continuant (/s,f,z,v/) and stop (/t,p,d,b/),
between constriction at the alveolar ridge behind the upper front teeth
(/s,z,t,d/) and constriction at the lips (/f,v,p,b/). Taken together these
eight phonemes, formed from three binary contrasts, constitute a little system
within the larger system of English phonology.

The particular selection of features used in any language is largely
determined by phonetic drift over time and by a complex of historical and
social forces. But the universal stock of phonetic features is presumably
constrained by human anatomy and physiology: It must be drawn from the (as
yet unspecified) intersection of what we can articulate with what we can
perceive. The goal of much work (e.g., Jakobson, Fant, & Halle, 1963; Chomsky
& Halle, 1968; Ladefoged, 1971) has been to define the smallest set of
universal features (perhaps fewer than 20) that will include all features that
Figure 1. A three-dimensional binary feature space, excerpted from the multidimensional feature space that describes the English phonological system.
may be distinctive in any language.

But there is more to the phonology of a language than the structure of its phonemic system. Each language also disposes of more-or-less elaborate rules for combining phonemes into words: These are the rules of its syllable structure. For example, in English the basic syllable structure can be represented as: \((C)(C)(C)V(C)(C)(C)(C)\), where \(C\) = consonant, \(V\) = vowel and parentheses indicate that the slot may or may not be filled. Thus, the simplest syllable is an isolated vowel. But in most syllables the required vowel is preceded by up to three consonants and followed by up to five consonants (the latter only in a few rare words such as "triumph'st") (Abercrombie, 1967).

Moreover, there are strict limits on the permissible consonant clusters. For example, in English, if two obstruents (stops or fricatives) occur together, the voicing of the second must match the voicing of the first. Accordingly, English words may begin with \(sp-, st-,\) or \(sk-\), but not with \(sb-, sd-\) or \(sg-\). Hence, too, the plurals in \(-s\) or \(-z\), (apes, lions), the present indicatives in \(-s\) or \(-z\) (she raps, she loves) and the past in \(-t\) or \(-d\) (rapped, loved). A subsidiary rule states that, if the two obstruents are formed by closure at roughly the same point in the vocal tract, a neutral vowel (the so-called schwa) must be inserted between them, giving the plural, roses, the present indicative, she kisses, the past, she hated. Most normal children, growing up among English speakers, have unconsciously learned these rules by the age of about six, and therefore have no difficulty in forming the correct plurals, presents and pasts of words they have never heard before (Berko, 1958).

The point of this example is to make clear that very much more is required to learn the sound structure of a language than the capacity to listen and to imitate. In fact, as we shall see below, even within its first year of life, the infant has begun to discover and apply rules.

THE FUNCTION OF PHONETIC FEATURES

We have defined features up to this point in articulatory terms. In part, this is because precise acoustic description, drawing on spectrographic analysis, has proved intractable. But it is principally because articulation is, in fact, prior to the acoustic signal. Indeed, it has been plausibly argued that the feature structure of spoken language was primarily a solution to the problem of getting high speed articulatory performance out of low speed articulatory machinery (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The feature structure permits a shift from one phoneme to the next by a change of no more than one or a few articulatory features. The value of articulatory ease is attested by the universal phenomenon of assimilation. Every language has many rules by which certain sounds or classes of sounds take on features of neighboring sounds, permitting a "lazier," and so more rapid, articulation. For example, the final \(n\) of the prefix syn- (synthesis, synecdoche) becomes \(m\) in symbiosis and sympathy, taking on the labial articulation of the following consonant. Similarly, normally voiced \(l\), sounded with laryngeal pulsing in light, takes on the voiceless feature of \(s\) in a word such as slight.
Of course, a gain for the speaker may be a loss for the listener. It is precisely such shifts in articulation and the consequent subtle shuffling of the acoustic properties of neighboring phonemes that have thwarted attempts at automatic speech recognition and given rise to the central problems for a theory of speech perception. Parallel (or co-) articulation of consonant and vowel in the integral ballistic gesture of the syllable (Stetson, 1952) gives rise to an acoustic signal in which the cues to a particular phoneme vary widely as a function of context and in which the boundaries between successive phonemes are obliterated. The tempting model that language might have been expected to offer for the division of motor behavior into "natural" units is thus a mirage. The units are not to be found either in the articulation or in the acoustic signal. The problem of segmentation appears to be solved by perceptual fiat. Not surprisingly, this has encouraged theorists of speech perception to invoke exotic perceptual mechanisms such as analysis-by-synthesis (Stevens & Halle, 1967; cf. Liberman et al., 1967) and "dedicated" property or feature detecting devices (see below).

Perhaps specialized perceptual mechanisms have indeed evolved to match the specialized motor mechanisms. There is strong evidence in vocal tract morphology, in tongue and lip innervation, in mechanisms for breath control during speech, and so on, that extensive adaptations for speaking did occur (Lenneberg, 1967; Lieberman, 1972; Du Brul, 1977). Perhaps these and matching perceptual adaptations (including specialized sensorimotor processes for imitation) underlie the evolution of language. However, once the capacity for language had evolved, man was able to deploy it in another mode. What is interesting is that, when he does so, as in American Sign Language, the formal structure of the system remains largely unchanged.

**AN ALTERNATIVE SIGNALING SYSTEM: MANUAL SIGN LANGUAGE**

Visual and tactile finger-spelling, like alphabetic and syllabic writing, are parasitic on speech: They simply transpose its units into another modality. However, some visual languages are independent of spoken language: for example, the sign languages of the American Plains Indians (West, 1960), of the Australian aborigines (Umiker-Sebeok & Sebeok, 1977), and of countless deaf communities in the various countries of the world (Stokoe, 1974). The signs of these languages do not necessarily correspond to the words of any particular spoken language, nor do the rules for their combination follow the syntax of any spoken language.

Consider, as an example, since it has been the most extensively studied, American Sign Language (ASL or Ameslan). Ameslan is a derivative of the French sign language introduced by Gallaudet to the U.S. in 1817: Users of Ameslan today are said to understand French SL better than British SL—evidence for the independence of sign and spoken languages. The first dictionary of Ameslan (Stokoe, Casterline, & Croneberg, 1965) contains over 2,000 signs. Many of them seem iconic, but usually not until one knows what they mean—just as one may not recognize the metaphor in, say, "The road runs west" until it is pointed out. Other signs are indexical: Pronouns, for example, are often formed by pointing. However, pointing and pure pantomime are rare. The overwhelming majority of signs are arbitrary or, if once iconic, have now lost much of their iconicity (Frishberg, 1975).
Signs may use one or two hands, and may vary along at least three orthogonal dimensions: configuration, position within the signing area (a rough circle around head and chest, centered below the chin), and movement. Stokoe et al. (1965) have analyzed the values along these dimensions into some 55 "chermes," a number well within the phonemic count of many spoken languages. Later work (e.g., Battison, 1974; Klima & Bellugi, 1979; Lane, Boyes-Braem, & Bellugi, 1976) has demonstrated that formational rules govern the possible combinations of chermes into signs, just as the phonological rules of a language govern the combination of phonemes into words. Finally, Amselan has now been shown to possess a richly inflected grammar and a syntax, that is, a set of rules governing the spatial and temporal ordering of signs into sentences (Klima & Bellugi, 1979; see also Siple, 1979). In short, Amselan displays all the distinctive properties of a human language including a dual pattern of form and syntax.

The significance of this recent work on Amselan is twofold. First, it underlines the link between hand and mouth, and the likely importance of a rapid, informationally-dense signaling system for efficient linguistic communication, a point to which we return below (see also Studdert-Kennedy, 1977). Second, it demonstrates the abstract nature of the capacities underlying language development. So far as we know, no other animal has developed a capacity for essentially equivalent communication in two different sensorimotor systems.

THE GREAT APES

Recent successes in training apes to communicate by means of artificial symbol systems (Premack, 1976; Rumbaugh, 1977) or a natural sign language (Amselan) (Gardner & Gardner, 1969, 1975; Terrace, Pettito, & Bever, 1976a, 1976b; Patterson, 1978) have shown that the cognitive, representational and perhaps even linguistic capacities of chimpanzees and gorillas, though vastly inferior, are nonetheless very much closer to those of man than was once thought. Given the tight genetic relation between man and chimpanzee (King & Wilson, 1975) and their very different ecologies, one may wonder whether these apparently similar behavioral capacities in man and ape may not be homologous capacities derived by genetic transmission from a common ancestor.

Unfortunately, the degree of similarity and its evolutionary implications are difficult to assess because none of the supposedly linguistic behaviors of the apes seems to occur naturally. All have required intervention by animals of another species in the form of systematic operant conditioning. This is particularly striking in the work of Premack (1976) and Rumbaugh (1977) where chains of behavior are established by direct shaping and primary reinforcement of hundreds of responses with food, drink, bodily contact and so on. For the signing chimpanzees, such as Washoe (Gardner & Gardner, 1975) and Nim (Terrace et al., 1976a, 1976b), the social reward of trainer approval is more usual. Nonetheless, even here the fundamental training procedure has been operant shaping and molding of specific behaviors. In other words, language learning in the great apes does not proceed without the establishment of stimulus-response contingencies.

By contrast, the human infant is apparently disposed to learn language even in the absence of specific response shaping and reinforcement. While it
too may require the generalized social reinforcement of a partner's attention, the infant does not require shaping and reinforcement of particular responses. On the contrary, as Brown (1973) has remarked, parents tend to reinforce the truth value, but not the form of their children's utterances. In other words, language develops in spite of the absence of narrowly-defined stimulus-response contingencies.

Particularly striking in this context is the recent work of Feldman, Goldin-Meadow, and Gleitman (1977) on the spontaneous development of signing in deaf children. They studied six deaf children, over an age span of 1,5 through 4,6 1/2, whose parents were following the "oralist" practice recommended by some authorities in the U.S.A. These authorities believe that signing to congenitally deaf children lowers their motivation to lip-read and articulate English; they therefore urge parents and siblings of such children to avoid all gestures, formal or informal. According to Feldman et al., the families of their six subjects were largely successful in following this practice.

The procedure of the study was to videotape each child playing and passing time with its mother and the experimenter during several standardized home visits. In the course of playing with the toys and games introduced by the experimenter every child devised its own "home-signs," that is, a characteristic set of motor-iconic gestures to refer to objects, actions, predicates. Moreover, each child gradually began to combine these signs into two-, three-, and even six-sign sequences, creating its own semantically-based syntax, including systematic deletion rules of the kind observed in a normal hearing child's "telegraphic" speech. This last point is particularly interesting, since telegraphic signing was not produced by the adults conversing with the children any more than is telegraphic speech under normal circumstances. The authors end their lengthy analysis with the conclusion that "...there are significant internal dispositions in humans that guide the language acquisition process" (Feldman et al., 1977).

There is, of course, no evidence for such dispositions in the ape. This argues that the cognitive capacities now being discovered in the apes are general rather than specifically linguistic. The adaptive functions of these capacities are not always obvious. For example, how does the wild chimpanzee use its capacity to symbolize? Or is this capacity perhaps a "neo-phytoype" (Kuo, 1976; Miller, 1980), an item of general behavioral plasticity, not normally deployed, but available for use in the face of the right selective pressures?

Another general capacity, impressively displayed in the recent language projects, does have obvious utility, namely, the capacity to learn a new motor response by observation and imitation. This requires that the animal, first, be able to parse perceived behavior into action components, and second, have sensorimotor connections by which the parsed patterns may be mapped into motor commands (cf. Terrace et al., 1976a, p. 21). Field observations attest to the role of imitation in the young chimpanzee's learning to fish for termites, for example, or to build its nest (van Lawick-Goodall, 1971).

Yet a third chimpanzee capacity, essential to linguistic communication, has recently been demonstrated by Premack and Woodruff (1978)--the attributing
of "intention" to the behavior of another organism. Here again, the capacity, whatever its linguistic worth, obviously contributes to the development of social intelligence. In fact, laboratory studies of ape "language acquisition" probably have more to teach us about the evolutionary origins of mind than of language. Certainly, as Limber (1977) suggests, conversational chimpanzees may offer an experimental approach to the study of relations between language and thought (for example, does naming facilitate problem-solving?), but the focus would then be on thought rather than on language. For insight into the origins of language, the frank analogues of birdsong may have more to offer than the possible homologues of ape signs.

**THE SONG BIRDS**

Unlike observational learning of other motor behavior, vocal learning can have no value beyond its use in communication. The analogous appearance of vocal learning in both man and bird is therefore of special interest (Marler, 1970, 1975; Nottebohm, 1970, 1975). Indeed, Marler has proposed as "...a significant evolutionary step toward...the strategy of speech development of Homo sapiens..."" the emergence of "...new sensory mechanisms for processing speech sounds..." as well as "...new neural circuitry...to modify patterns of motor outflow so that sounds generated can be matched to preestablished auditory templates" (Marler, 1975, pp. 32-33). As we shall see, the evidence for "new sensory mechanisms" or "auditory templates" in humans is weak, but there is good evidence for specialized sensorimotor processes.

**Templates and Feature Detectors**

Birds (and other animals). Species-specific templates were proposed by Marler (1963, p. 233) and Konishi (1965) to account for the fact that many songbirds prefer to learn the songs of their own species. Even if they are deprived of conspecific song during the sensitive phase, and are exposed to the songs of closely related species, they tend not to learn them (e.g., Marler & Peters, 1977).

The form of these templates, "...lying in the auditory pathway" (Marler, 1975, p. 26) has never been specified. However, presumably they would consist of networks of specialized neurons tuned to particular properties of the species' song. Cortical neurons sensitive to changing frequencies were reported for the cat ("miaow cells") by Whitfield and Evans (1965). Cells tuned to species calls have been reported for the bullfrog (Frischkopf & Goldstein, 1963; Capranica, 1965), the squirrel monkey (Wollberg & Newman, 1972), several species of echo-locating bat (Neuweiler, 1977) and the starling (Leppelsack & Vogt, 1976).

Humans. A possible analogy between species-specific call or song detectors and phonetically relevant, acoustic feature detectors was not lost on students of speech perception (e.g., Abbs & Sussman, 1971; Liberman et al., 1967; Studdert-Kennedy, 1974). The feature detector promised to solve at a single blow a variety of problems in speech perception, including that of syllable segmentation. Moreover, the notion of feature with its roots in ethology, linguistics and pattern recognition was attractive to biologically-inclined students of language, looking for signs of an innate acquisition device (e.g., Stevens, 1975). Unfortunately, the several lines of evidence
and speculation seem to have converged on an error.

The story begins with the phenomenon of categorical perception (Eimas, 1963; Liberman et al., 1967; Studdert-Kennedy et al., 1970). Early work with speech synthesizers showed that it was a simple matter to construct acoustic tokens of opponent phonetic types by varying a single acoustic parameter. For example, by varying the interval between plosive release and the onset of laryngeal pulsing, that is, voice onset time (VOT), one could construct a continuum of, say, a dozen tokens ranging in equal acoustic steps from /ba/ to /pa/, or from /da/ to /ta/.

If listeners were asked to identify these tokens, they showed a strong tendency to call any particular stimulus by the same name (e.g., /ba/) every time they heard it. There were few, if any, ambiguous tokens. Furthermore, if they were asked to discriminate between neighboring pairs of tokens, they tended to do badly if they judged the two tokens to be members of the same phoneme class, but well, if they judged the tokens to be members of opponent phoneme classes—even though the acoustic interval between pairs was identical in the two cases. This phenomenon, dubbed "categorical perception," seemed to be a useful process for speech perception. After all, one cannot afford to judge a word to be more-or-less "bat" or more-or-less "pat." One must categorize it instantly as one or the other: Classification is a crucial process in phonetic perception.

The next event in the story was the demonstration by Eimas, Siqueland, Jusczyk, and Vigorito (1971), using a non-nutritive sucking habituation procedure, that one-month- and four-month-old infants could discriminate between two tokens differing by 20 msec along a voice onset time continuum, providing they were tokens that adults normally classified as different phonemes. But the infants could not discriminate between tokens that adults normally classified as the same phoneme. Similar results for a variety of synthetic speech continua were reported in due course for infants growing up in other language communities (see Eimas, 1975, for a review).

The suspicion that these results reflected categorical perception, mediated by specially tuned, innate feature detectors, was not easy to resist—particularly since the phylogenetic emergence of such detectors might then be the evolutionary step that carried hominids from a graded to a categorical communication system (cf. Marler, 1975). The hunt for independent evidence of such detectors operating in human adults began, and, by 1973, Eimas and Corbit were able to report apparent success. They modified a procedure with a long history in visual studies: adaptation. The paradigm is simple enough. For example, prolonged fixation of a red patch of light adapts or fatigues a red detector cell and relatively sensitizes its opponent green detector cell, so that upon looking at a white screen, the viewer sees a relatively unsaturated green patch the same shape as the red adaptor. Related effects in form and tilt also occur. Such effects have frequently been taken as evidence for the operation of opponent feature detectors.

Eimas and Corbit (1973) asked listeners to categorize members of a synthetic voice onset time continuum and demonstrated that the perceptual boundary between voiced and voiceless categories along that continuum was shifted by repeated exposure to (that is, adaptation with) either of the
endpoint stimuli: There was a decrease in the frequency with which stimuli close to the original boundary were assigned to the adapted category and a consequent shift of the boundary toward the adapted stimulus. They took the effect to be evidence for the operation of an opponent feature detecting system. Several dozen studies over the next five years replicated the effect on several other synthetic speech continua. (For reviews, see Ades, 1976; Cooper, 1975; Eimas & Miller, 1978.)

Thus, the chain of inference and speculation from percept to detector was complete. Unfortunately, each link in the chain has proved weak. First, several studies have shown that categorical perception is not peculiar to speech, or even to audition. For example, Pastore et al. (1977) demonstrated categorical perception of critical flicker, with a sharp boundary at the flicker-fusion threshold. Second, other studies (e.g., Carney, Widin, & Viermeister, 1977) have demonstrated that the degree of categorical perception varies with the experimental method used to measure it: Listeners can be trained to hear a supposedly categorical continuum noncategorically or to shift category boundaries from one point on a VOT continuum to another. Finally, cross-language studies have found that speakers of different languages may place phonetic boundaries at different points along the same acoustic continuum, demonstrating that acoustic-phonetic categories are determined linguistically by language experience rather than neurophysiologically by innate feature-detecting devices. (For a review of cross-language studies, see Strange & Jenkins, 1978.)

The demise of categorical perception as a specialized phonetic process also cuts the only links in the chain. Thus, instances of what appears to be infant categorical perception will doubtless find a straightforward account in terms of auditory psychophysics, similar to that developed for the adult case. In fact, Pisoni (1977) has already developed such an account for voice onset time.

By the same token, we no longer need opponent process feature detectors to account for a general psychophysical phenomenon—particularly since there are quite other grounds for doubting the opponent detector model. Most obvious is the model's lack of behavioral or neurological motivation. For, while the facts of additive color mixture and retinal neurophysiology make an opponent detector account of after-effects entirely plausible, the facts of perceived stop consonant onset and cochlear neurophysiology certainly do not. However, an adequate discussion of speech adaptation is well beyond the scope of this chapter, and it must suffice to remark that plausible accounts of the effects in terms of stimulus range (Rosen, in press), auditory contrast (Simon & Studdert-Kennedy, 1978) or other more general processes (Remez, 1979) have already begun to appear.

We must conclude that we now have no evidence for the operation in speech perception of specialized sensory mechanisms analogous to the auditory templates postulated for certain songbirds.

**Lateralization and the Sensorimotor Device**

**Birds.** One of the most remarkable discoveries in recent years is the lateralization of neural function in birdsong (Nottebohm, 1971, 1972, 1977)—
at present, the only securely attested instance of lateralized behavior outside man (although see Dewson, 1977, and Petersen, Beecher, Zoloth, Moody, & Stebbins, 1978). The typical songbird syrinx, as instanced by that of the canary (Nottebohm, 1977), has two independently innervated and functionally separate halves. Sections of the right and left halves (or of their innervating hypoglossal branches) have very different effects: Right side sections lead to the loss of no more than 0 to 15% of pre-operative song syllables, while left side sections lead to a 90 to 100% loss. Similar effects of peripheral lesion have been observed in the chaffinch (Nottebohm, 1971, 1972) and the white-crowned sparrow (Nottebohm & Nottebohm, 1976). For the canary, Nottebohm (1977) has also traced motor pathways from the syrinx to the associated brain structures: Unilateral brain lesions indicate that the left hemisphere contributes radically more to song control than does the right.

All these effects are motor, and no perceptual lateralization has been demonstrated. However, it is of interest that the principal motor control center lies next to the telencephalic auditory projection, where processes involved in establishing the species-specific song template are believed to occur. Indeed, it was Nottebohm's (1970) original notion that lateralization might be associated with complex learned behavior. This view has been thrown into question by the discovery of peripheral lateral equipotentiality in the orange-winged Amazon parrot (Nottebohm, 1976), a bird well-known for its vocal plasticity, and of left lateralization in the domestic fowl (Youngren, Peek, & Phillips, 1974), a bird of equally well-known vocal stereotypy. Nonetheless, current research on the canary is charting neural links between the two centers, in an attempt to isolate the sensorimotor connection, presumably essential to song learning (Kelley & Nottebohm, 1979).

Humans. It has been known for many years that the left cerebral hemisphere contributes more to language function than the right, in most normal humans. The bulk of our knowledge comes from studies of aphasia, induced by stroke, tumor or gunshot wound (e.g., Jenkins, Jimenez-Pabon, Shaw, & Sefer, 1975; Penfield & Roberts, 1959) and, more recently, from studies of "split-brain" patients, whose cerebral hemispheres have been surgically separated by section of the connecting pathways for relief of epilepsy (e.g., Zaidel, 1978a, 1978b). The latter condition permits an investigator to assess the linguistic capacities of each hemisphere independently.

Of particular interest, in light of the bird song findings, is that left hemisphere specialization seems to be primarily for control of the articulatory apparatus and for perceptual analysis of spoken words into their phonetic segments. The human larynx and its associated articulatory structures (tongue, velum, jaw) are bilaterally innervated, but unilaterally controlled. Thus, "verbal apraxia," or aphasic disturbance of articulation, is associated with damage to motor areas of the left hemisphere. By corollary, the right hemisphere, despite a fair capacity for understanding speech, is essentially (that is, apart from a limited capacity for expletive and non-propositional utterance) mute. Interestingly, skilled manual movements (Kimura & Archibald, 1974) and non-verbal oral movements (Mateer & Kimura, 1977) tend also to be impaired in cases of non-fluent aphasia. Moreover, disturbances of sign-language in the deaf are associated with left-hemisphere damage (Kimura, Battison, & Lubert, 1976). After a review of such
evidence, Kimura (1976) suggests that "...the left hemisphere is particularly well adapted, not for symbolic function per se, but for the execution of some categories of motor activity which happened to lend themselves readily to communication" (Kimura, 1976, p. 154).

However, more than motor specialization is involved. Studdert-Kennedy and Shankweiler (1970) concluded, from a study of normal subjects' performance on a test in which competing nonsense syllables were presented simultaneously to left and right ears, that the left hemisphere was specialized for phonological analysis of spoken language. Recent work with split-brain patients has confirmed this conclusion (Zaidel, 1978a). The dissociated right hemisphere of such a patient has a sizeable auditory lexicon and a rudimentary syntax sufficient for understanding phrases of up to three or four words in length. However, it is incapable of identifying nonsense syllables or of recognizing that, say, "rose" rhymes with "toes" (Levy, 1974). In other words, the right hemisphere is not only mute, but is organized by meaning rather than by linguistic structure. Unlike the left hemisphere, it perceives language holistically, seizing meaning from the "auditory contours" of words rather than by phonological analysis. If, as we suggested earlier, the characteristic feature structure of speech sounds derives from articulatory constraints, we should perhaps not be surprised to discover that their perception is linked neurologically to their production.

Direct evidence for a sensorimotor link in the left hemisphere comes from the work of Sussman (1970; Sussman & MacNeilage, 1975; Sussman & Westbury, 1978). Sussman devised a bizarre tracking task in which a sinusoidal waveform, fed into one ear, can be tracked (i.e., copied) by movements of tongue, jaw, lips, or hand. The results of the tracking movements, electronically multiplied into the audio-frequency range, are then fed to the opposite ear. In several experiments, Sussman and his colleagues have shown that tracking movements made by a speech articulator (tongue, jaw, lips) are more accurate if auditory feedback from the movements comes to the right ear (i.e., left hemisphere) rather than to the left. In all but one of the control experiments in which tracking movements were made by hand, there was no ear difference. Sussman and MacNeilage (1975) concluded that their results reflected "a lateralized, speech-related, auditory-sensorimotor integration mechanism" (1975, p. 139).

The ultimate function of such a mechanism is, of course, unknown. However, if anything is to be made of the analogy with bird song, we may speculate that unilateral control is necessary for motor coordination of a bilaterally innervated apparatus (cf. Levy, 1969; Liberman, 1974; Marler, 1970). This might be achieved either by assigning execution primarily to one side of the peripheral apparatus and therefore to lateralized control centers in the brain (as seems to be done in the canary) or by assigning to one side of the brain central coordination of a symmetrically innervated peripheral apparatus (as seems to be done in the human). Lateralization of the associated perceptual center would then follow to facilitate sensorimotor learning. In the human case, evolution of the sensorimotor mechanism led further to development of a lateralized syntactic device, itself perhaps motoric in origin and specialized for precise, temporal coordination of hierarchically ordered structures. The result is that the left hemisphere "...does seem to possess an innate and highly specialized linguistic mechanism
whose paradigmatic functions are phonetic and syntactic encoding and analysis" (Zaidel, 1978a, p. 196).

Finally, in the human case, lateralized control of the vocal apparatus seems to have been laid down on the neural substrate of manual lateralization, already evolved for tool use and/or gestural communication (Levy, 1974). Semmes (1968) has provided an account of the association by arguing (from a lengthy series of gunshot lesions) that the left hemisphere is focally organized for fine motor control, the right hemisphere diffusely organized for broader control. More generally, Zaidel (1978b) has suggested that "...each hemisphere specializes for a different style of information processing..." (p. 263), and Levy (1974) proposes that hemispheric specialization may achieve functional dissociation of neurologically incompatible behaviors. But the important point here is not the possible complementary functions of the cerebral hemispheres (Zangwill, 1960). Rather it is the notion, developed by Kimura (1976) and touched on in our discussion of manual sign language, that the origin of cerebral lateralization for language is in the control of skilled movement rather than in any "higher" symbolic processes.

What is puzzling, of course, is that, unlike song lateralization in birds, which has been observed in virtually every individual studied (Nottebohm, 1977), human lateral specializations are neither uniform across the population nor perfectly associated. The incidence of right-handedness in the U.S. population is estimated at roughly 90% (Hardyck & Petrino-vich, 1977; Levy, 1974), and the incidence of left dominance for language at roughly 95% among the right-handed, 60% among the left-handed (Milner, Branch, & Rasmussen, 1964). If such figures prove reliable across the human population, the network of lateralized functions would seem to offer an instance of an "evolutionarily stable strategy" (Maynard-Smith & Price, 1973), a balanced polymorphism that it will be a challenge to explain.

Sensitive Phases

Birds. Many songbirds can only learn their species' song if they are exposed to that song during a sensitive phase. The phase may range from as little as 40 days for the white-crowned sparrow through 10 months for the chaffinch, to as long as two years for the Oregon junco (Petrinovich, 1972). In some birds, such as the white-crowned sparrow or the marsh wren, there may be two distinct phases, separated by weeks or even months: an input phase for perceptual learning and an output phase for subsong and learning to sing. In other birds, such as the chaffinch, the two phases may overlap, with elements of subsong appearing before the input phase has ended. Presumably such variations have adaptive value and can be related to the ecologies and life-histories of the different species. In fact, Immelmann and Suomi (1980) point out that it is precisely the systematic variations across species in temporal patterns of song-learning that validate the concept of a sensitive phase and prove it to be more than a handy descriptive term for a process begun by maturation and ended by song acquisition. Much recent work is therefore aimed at pinning down the ultimate selective pressures (Kroodsma, 1980).

However, the proximate mechanisms controlling onset and offset of sensitive phases are not well understood. Hormone levels are often suggested (e.g., Bateson, 1973). Nottebohm (1967) castrated a male chaffinch during its
first winter, thus precluding either the learning or the singing of song during its first spring. In the second spring the bird was implanted with a testosterone pellet and proved able to learn two tutor songs, but no more. Nottebohm suggests that "...the ability to develop song for the first time is not age-dependent" (1967, p. 278). However, "age" is a cover term for aspects of physical maturation as well as for the mere passage of time. Since castration may have delayed, if not halted, normal maturational processes, the experiment does not rule out physical maturation as the determinant of the onset of song-learning. Since, moreover, a total of two songs falls within the normal chaffinch repertoire range of 2 to 6 songs (Nottebohm, 1967), we might reasonably hypothesize that song-learning had ceased when the available "neural space" was filled (cf. Bateson, in press; Kroodsma, in press). The point here is that, as Immelmann and Suomi (1980) remark, specialized proximate mechanisms beyond physical maturation and neural preemption may not always be necessary for delimitation of a sensitive phase in song birds.

Humans. Lenneberg (1967, pp. 125-187) was the first to postulate a "critical period" for language learning. He was careful to make clear that he was offering no more than an analogy with the critical periods (or sensitive phases) of filial imprinting and song learning in birds. He places the period roughly between the end of the second year and the beginning of the twelfth. Broadly, his argument is based on: (1) the regularity of the time of onset of speech across cultures; (2) the different effects on language of various pathologies, particularly cerebral insult and deafness, as a function of age: in general, the younger the child at the time of brain injury or the older at the time of onset of deafness, the better the prognosis for language development; (3) the commonly observed, increased difficulty of learning a foreign language after puberty—at least without appreciable interference from already known languages. Within the critical period, Lenneberg argued, languages are fully learned by mere exposure; after the critical period they are learned less well and with increasing difficulty—an analogy with song-learning in the zebra finch (Immelmann, 1969).

Lenneberg attributes onset of the "critical period" to general maturation of the central nervous system. Cerebral structure (cell density, dendritic arborization) and chemical composition, as well as characteristic brain wave rhythms measured by electroencephalography, have reached roughly 75 percent of their adult asymptotic values by the age of two years. Thus, Lenneberg does not propose, nor is there any evidence for, a specialized onset mechanism analogous to the changes in hormone levels postulated for some birds.

The lateness of the proposed onset is largely a matter of definition. Since Lenneberg regarded syntax as the distinctive property of language, he identified language onset with the first putting together of words. This typically occurs between 18 and 28 months. Moreover, Lenneberg specifically denied the importance of experience during the first two years, largely on the grounds that children deafened as late as the end of their second year find it no easier to learn language than do those who have been deaf since birth. However, his evidence for this is drawn entirely from informal personal observation, and it seems unlikely that the orderly progression during the first year of life from prespeech oral play through cooing, intonation and babbling is devoid of functional value. If we take the presence of language-specific structure in infant babble at roughly 8 months (Mehler, personal
communication), or even the prespeech lip and tongue movements in train with a mother's behavior (Trevorthen, Hubley, & Sheenan, 1975), as evidence that language sensitivity has begun, we may place the onset of the sensitive phase in the second half of the first year or even as early as the second month of life. The factors controlling this onset may still then be, as Lenneberg proposed, a combination of physical maturation and appropriate environmental stimulation.

The difficulty of learning a language after puberty is commonly known. Formal evidence for the likelihood of both grammatical and articulatory defects in a second language learned as an adult comes from Oyama (1973, cited by Krashen, 1975). Evidence for even greater defects in a first language learned after puberty has recently come from Genie, a California "wild child" (Curtiss, 1977). When discovered at the age of 13 1/2 years, after nearly twelve years of brutal undernourishment and isolation in a silent back room, Genie had virtually no language. Five years later, she had learned some language by "mere exposure" without specific training. Interestingly, her capacity for phonetic perception was normal, perhaps because her isolation had not begun until 20 months, when the phonetic groundwork had already been laid and she had begun to speak a few words. But her speech was severely distorted and her syntax deficient—for example, she could not use any wh- question words, verbal auxiliaries or embedded structures. In other words, she learned language very much less well than a normal child, as Lenneberg would have predicted.

The factors controlling offset at puberty are not known. Lenneberg proposed loss of cerebral plasticity due to completed lateralization of function—without, however, offering any suggestion as to why language should be lateralized. His argument was based on clinical evidence of recovery from aphasia as a function of age. The picture has been confused by recent work suggesting that lateralization may be present from birth (Molfese, Freeman, & Palermo, 1975; Entus, 1977; Glanville, Best, & Levenson, 1977), and essentially complete by five years—roughly coinciding with the time when first language acquisition is approaching completion (Krashen, 1975). But the question of offset mechanism is important if the concept of a sensitive phase for language learning is to retain validity.

The reason for this is that we cannot justify the concept by referring to inter-species differences of the kinds observed in song birds, nor by reference to its onset mechanism, since this appears to correlate with general physical maturation. If, further, its offset mechanism were merely preemption of "neural space," as the articulatory, syntactic and even lexical interference between earlier and later learned languages perhaps suggests, we might be dealing with a general loss in cerebral plasticity and with a process common to other classes of behavior rather than one peculiar to language. In short, the validity of the concept may rest on demonstrating that the offset mechanism is directed specifically at language learning. At present we have no evidence that this is so.

Finally, we must ask what the function of a sensitive phase for language might be (cf. Bateson, in press). First, following Immelmann (1976, p. 152), we must distinguish between the period during which a behavior can be learned and the period during which it normally is learned. It is on the offset of
the former that we might expect selective pressures to bear. If offset were early, roughly contemporaneous with release of offspring into a peer world, the language learned would be that of the parents, and we might reasonably suspect that a sensitive phase ensures a dialect that will attract sexual partners from ecologically similar backgrounds. Dialects might then, indeed, be "signs of incipient speciation" (Marler, 1963, p. 796; cf. Armstrong, 1963, chap. 5). Such a function is unlikely in humans, despite the presumably high correlation between inbreeding and dialect in, say, the highlands of New Guinea or of Austria, because many more salient features (such as habitat and body ornament) serve to isolate human breeding populations.

Moreover, offset in humans is relatively late, well beyond the point where the child has abandoned the nuclear family for its peers. Accordingly, whether a child learns the dialect of its parents rather than of its peers (as is said of some English upper-class children thrown, by the accidents of war, among lower-class peers), or of its peers rather than of its parents (as do the children of non-English-speaking immigrants to Australia or the U.S.A.), may sometimes depend on social rather than directly biological factors. An echo in the behavior of Bewick's wren, which learns the song not of its father but of neighbors in its newly chosen breeding site (Kroodsma, 1974), suggests that social bonding may be among the biological functions of dialects in both bird song and language (cf. Petrinovich, 1972).

Whether this function is important enough to bear the weight of accounting for a sensitive phase in language learning, one may doubt. In fact, given the weakness of this function and the lack of any clear evidence for proximate controlling mechanisms directed specifically at language, one may be tempted to conclude that a "critical period" for human language acquisition is more apparent than real, a mere matter of cerebral maturation in its onset and of neural preemption (or, as in the case of Genie, atrophy) in its offset.

THE INFANT AS PATTERN SEEKER

In songbirds, both species-specific template and sensitive phase are adapted to the same end, namely, acquisition of the species song within a few months of birth. The song to be learned is generally brief and simple. A template ensures that from the varied songs around it, the young bird will learn to recognize (if female) as well as to practice and execute (if male) the song of its own species, while a sensitive phase usually confines learning to the weeks before dispersal from the home site and/or to the months after the bird has settled among its breeding peers. Nonetheless, not all birds that learn to sing have either a template or a sensitive phase. Indeed, certain mimics, such as the North American mockingbird, learn, presumably without template and even late in life, the songs of species quite unrelated to themselves. Perhaps it is among such generalized, all-purpose song learners that we should look for an analogy with the human infant.

In any event, far from being constrained to learn the sound pattern of its language within a few months of birth, the human newborn has before it some two years of infancy. Moreover, what it must learn is not merely to imitate the sounds of the speakers around it—important though this undoubtedly is—but also to perceive and deploy their characteristic sound system. Rather than narrowly defined templates we might therefore expect the infant--
and its caretakers—to have evolved broad behavioral programs that will encourage vocal interchange and facilitate discovery of spoken pattern. The general process of acquisition seems, in fact, to be one of gradual differentiation: sound from silence, voice from sound, mother's voice from stranger's, intonation from monotone, syllabic beat from intonated melody, consonant from vowel, perhaps feature from phoneme.

One-day-old infants will suck a pacifier to turn on music and soon begin to prefer voices to music (Friedlander, 1970). Indeed, within a few days of birth, breast-fed babies have learned to turn toward a voice, twisting the mouth as if in expectation of a nipple and crying when none is there (Alegría & Noirot, Note 1). By 20 to 30 days the infant has learned to recognize its mother's voice, as she reads from behind a screen, and will suck more rapidly for her voice than for a stranger's (Mills & Melhuish, 1974)—provided that she speaks with her customary intonation rather than reads backwards from a text (Meehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978).

From around the second month, the infant becomes accessible to "conversations" with its mother, watching her eyes (humans are the only animals with permanently visible whites to their eyes, contrasting with the iris), smiling, moving lips and tongue in apparent imitation of the mother ("prespeech") and gurgling (Trevorthen et al., 1975). With the child's discovery that events in the external world—particularly, the vocalizations, touches, gestures of its mother—may be contingent on its own behavior, the way is opened for games (e.g., "peekaboo"), rhythmic interactions, cooing and laughter (Watson, 1977; Papousek & Papousek, 1975). The very precise temporal patterning of mother-infant interaction, with its alternating vocalizations, pauses, exaggerated facial displays, and so on, lays the ground for later social interchange (Stern, Jaffe, Beebe, & Bennett, 1975). Freedle and Lewis (1977) find that vocalization occupies a special place in early mother-infant interaction: It is more likely to accompany playing, looking, holding or touching than changing, feeding or rocking. Moreover, vocalization by one partner is the most likely behavior to follow vocalization by the other, leading to the conclusion that "...vocalization is the central behavior which maintains interaction" (Freedle & Lewis, 1977, p. 160). However, this interactive pattern is not specific to the vocal modality: For deaf children, growing up as signers, signing occupies the privileged position (Feldman et al., 1977). From this we may conclude that mother-infant interaction is broadly adapted to the development, not simply of speech, but of any communicatively viable signaling system. This, in turn, suggests that the infant's discovery of speech may be guided by the pattern of input from its environment rather than by the triggering of tuned detectors.

Of interest here is the nature of the mother's vocalizations, that is, of what has come to be called "baby talk" (BT), the style of speech used by adults, and even young children, when addressing infants (as well as animals and lovers). Baby talk has been studied in many cultures and is characterized by what Ferguson (1978) has termed a "simplified register." The principal acoustic characteristics of this register are, according to Sachs (1978), that it has an overall higher pitch, a wider frequency and intensity range and a more markedly regular rhythmic structure (cf. nursery rhymes). In short, BT exaggerates the acoustic contrasts on which speech is based. While it is unlikely that any single property of the speech addressed to the infant is
essential to normal development (cf. Newport, Gleitman, H., & Gleitman, L. R., 1978), it is equally unlikely that a culturally widespread phenomenon such as BT is devoid of function. If function can be inferred from structure, the function of BT is to draw the infant's attention to important acoustic contrasts in speech (cf. Garnica, 1978) and to launch it on its search for pattern. Thus, we may see BT as the exogenous auditory counterpart of the endogenously controlled eye movements and head turning with which the human newborn searches for visual contour (Haith, 1978).

What the infant has learned perceptually about its native language begins to emerge in babble, around the sixth-to-ninth month. Jakobson (1968) dismisses babble as irrelevant to language acquisition on the grounds that it is primarily a motor activity, devoid of linguistic import. He is correct, inasmuch as normal perception of speech and language, as well as a highly educated level of reading and writing, can be developed, by prolonged and careful instruction, even when articulation has been pathologically precluded since birth (e.g., Fourcin, 1975). But this does not mean that, under normal circumstances, babbling contributes nothing to perceptual or, especially, expressive development. Indeed, it is unlikely that a behavior so regular in its time of onset and developmental course should altogether lack function.

Babble offers an obvious analogy with subsong, the low-intensity, "generalized" singing that precedes true song in many songbirds. Here, too, function is in doubt, because subsong tends to recur each year, as though it might simply reflect lower motivation in early Spring or late Fall (Thorpe, 1956, p. 373). Moreover, the female learns to recognize the male's song even though she herself (like the pathological human cases cited above) never engages in subsong. Nonetheless, subsong does last longer in the bird's first year and bears several interesting analogies with babble—enough to suggest that both activities may be necessary to normal motor, if not to normal perceptual, development.

In the chaffinch, for example, subsong seems to be a poorly differentiated version of the species song with a much greater frequency range. Learning involves dropping unwanted elements and organizing the remaining notes into the correct rhythm (Thorpe, 1956, p. 374), presumably to accord with the inborn template, as modified during early months of the sensitive phase. In the human, babble also seems to begin as a poorly differentiated stream, with many more components than will eventually be used. Gradually, over the course of two or three months, the stream begins to take on properties of the native language, presumably revealing what the infant learned perceptually during its first months of life. Just what these properties are is not yet known, partly because reliable phonetic transcription is difficult. Intonation is the most obvious property, and characteristic pitch contours can be traced in spectrograms (e.g., Nakazima, 1962), but language specific consonant-vowel syllables may be present also (Nakazima, 1975; Kewley-Port & Preston, 1974; Huxley & Ingram, 1971, pp. 162 ff.). In any event, Mehler (personal communication) reports that French-speaking adults can reliably identify infant babble, even in the second month of babbling, as French or not-French.

All this is consistent with the view that babble and subsong enable the organism to discover the limits of its vocal apparatus and to establish necessary sensorimotor links. Here, however, parallels between bird and
infant cease. For while the end of subsong is true song, of which the use
does not have to be learned, the end of babble is merely a modest articulatory
repertoire, already language-specific, but enough for no more than a start on
the discovery of a linguistic system.

The process of discovery is, so far as we know, without parallel in the
communication system of any other animal. The infant does not simply imitate,
matching a particular utterance to a particular type of situation. Rather, it
searches out contrasts among components of its own repertoire and uses them to
signal contrasts in its desires, experience or behavior. Often, the con-
trasts, in both signal and message, are entirely novel and without counterpart
in the adult system.

The process is well illustrated in a recent study by Menn (1979). She
followed the development of intonation (pitch contour) in the babble and early
speech of an American English boy between the ages of about thirteen and
fifteen months. She classified his behavioral routines into categories, such as
greeting, curiosity, narrative, desiderative, donative. Then she classi-
fied the pitch levels of babble in these situations as either moderate or
high, and the pitch contours as either rising or falling. Finally, she
correlated pitch levels and contours with behavioral routines.

Among the outcomes, predicted from adult speech and observed in the data,
were that "narrative" routines were accompanied by falling contours, while
"curiosity" or "desiderative" routines were almost always accompanied by
rising contours. However, the most interesting finding was that rising
"desiderative" contours, addressed to adults, were split according to pitch
levels into high (peak above 550 Hz) and moderate (peak below 450 Hz),
according to whether the child was seeking an object (e.g., food, toy) or
social interaction (e.g., play). In other words, at a stage of his linguistic
development when isolated words were still rare and word combinations did not
occur at all, this boy had constructed a sub-classification of his own rising
pitch contours into "moderate" for sociable occasions and "high" for object-
seeking occasions. Since, as Menn (1979) points out, adult speakers of
American English do not reliably use absolute pitch to contrast the uses they
wish to make of other people, we must conclude that the child had created its
own "erroneous" rules of intonation.

Such invention is not without precursor. The process of discovering
meaning, and of seeking its correlates in the gestures or vocalizations of
others, probably begins with the earliest mother-infant interchanges
(cf. MacNamara, 1972; Bruner, 1975). In due course, the infant chances upon
such correlates in its own or others' vocal repertoires and, with recognition
of the first contrasts in intonation, there begins the slow discovery of sound
pattern that will end, several years later, in a full and intricate phonologi-
cal system. For this and for the parallel processes of syntactic development
we find no analogues among birds or apes.

CONCLUSIONS AND QUESTIONS

---A language is an open system, adapted by its dual structure of sound
pattern and syntax for unlimited communication. If, as was argued, the dual
structure evolved to interface man's intellect with his peripheral anatomy,
it is unlikely that analogous duality of patterning will be found in animals of appreciably lower cognitive complexity.

—A dual structure is also found in manual sign language. That sign languages are manual emphasizes the importance of rapid articulatory gestures to effective linguistic communication. That they also display duality of patterning demonstrates the abstract nature of linguistic capacity: So far as we know, no other animal has developed two essentially equivalent systems of communication using different sensorimotor systems.

—Since none of the supposedly linguistic behaviors of the great apes occurs in a natural environment, recent successes in training them to communicate symbolically have little bearing on the origins of language. However, laboratory studies of the apes may lend insight into the evolution of intelligence and into relations between language and thought.

—Since the capacity for vocal learning has no value beyond its use in communication, its appearance (and pivotal social role) in both man and songbird is of great interest. However, of several possible analogies between birdsong and language learning—auditory templates, sensitive phases and lateralized sensorimotor mechanisms—only the last invites fruitful speculation. Lateralized motor control of birdsong, as well as the association of speech, right-handedness and manual sign language with left hemisphere mechanisms in humans, suggest that the origin of cerebral lateralization for language may be in the control of skilled sequences of movement. Future work might profitably explore functional relations among manual skills and the perception and production of both speech and sign language, in an attempt to establish the nature and extent of neural overlap.

—The long period of human infancy, taken with the diversity of human languages (both spoken and signed), suggests that biological adaptations for language learning are likely to be behavioral rather than tightly neurophysiological. Study of these behavioral adaptations, particularly of mother-infant interaction during the first year of life, may bring fuller understanding of language and of how it is learned.

REFERENCE NOTE


REFERENCES

Kelly, D. B., & Nottebohm, F. Projections of a telencephalic auditory


Levy, J. Psychobiological implications of bilateral asymmetry. In
Mills, M., & Melhuish, E. Recognition of mother's voice in early infancy.
Nature, 1974, 252, 123.


Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B., & Stebbins,


Studdert-Kennedy, M. *Speech perception*. In N. J. Lass (Ed.), *Contemporary


Terrace, H. S., Petitto, L., & Bever, T. G. Project Nim: Progress report II. New York: Department of Psychology, Columbia University, 1976. (b)


Zaidel, E. Lexical organization in the right hemisphere. In P. A. Buser and A. Rougeul-Buser (Eds.), Cerebral correlates of conscious experience. Amsterdam: Elsevier/North-Holland Biomedical Press, 1978, 177-197. (a)
