Language Development from an Evolutionary Perspective*

Michael Studdert-Kennedy†

EVOLUTION AND DEVELOPMENT
Preliminary

In his famous course of lectures at the University of Geneva (1906-1911), de Saussure distinguished langue, language as a system, a cultural institution, from parole, language as spoken and heard by individuals: "...language is not complete in any speaker; it exists only within a collectivity...only by virtue of a sort of contract signed by members of a community" (de Saussure, 1966, p. 14). Language was thus seen as an abstract property of a group, related to its variable individual speakers somewhat as a species is to its variable individual members.

Nineteenth century attempts to apply evolutionary principles to language did indeed view each language as a species. For example, Darwin wrote: "The formation of different languages and of distinct species...[is] curiously parallel...We find in distinct languages striking homologies due to community of descent, and analogies due to a similar process of formation" (1871, pp. 465-466). Here, Darwin was following in the steps of August Schleicher (1821-1868), who drew on the earlier work of Darwin himself and on the vast scholarship of 19th century European philology, to construct a taxonomy of Indo-European languages. Later work has continued to draw on Darwinian principles to construct evolutionary trees of language families (Lehmann, 1973).

However, a strict analogy between languages and species is untenable for both linguistic and biological reasons. For the historical linguist the most obvious difficulty is that a biological species is a reproductively isolated population: properties of one species cannot pass to another. Yet languages in contact clearly do influence one another, even after they have become discrete sociopolitical entities. Often we cannot then know how far shared properties have arisen from contact, how far from common descent.

A more serious difficulty follows from two central tenets of neo-Darwinian theory (the modern synthesis of Mendelian genetics with Darwinian natural selection). First, in the evolution of a species, the principal unit of variation and selection is not the species, but the individual organism. Individuals within a species differ in the number of offspring they produce; genes whose expression increases the relative number of an individual's offspring will increase their own relative number in the species' gene pool. The characteristics of a species are therefore determined by competition among individuals of that species. And, by analogy, the characteristics of a language are formed by competition among speakers of that language. So far, so good.

The difficulty arises when we combine the principle of individual selection with the so-called central dogma of modern biology, the "Weismann barrier," insulating germ cells from body cells: genes alter body cells, but body cells cannot alter genes. In other words, biological evolution does not proceed by the transmission of acquired characters across generations, and this is precisely what an evolutionary model of language change requires. We must therefore distinguish the cultural, or Lamarckian, evolution of language, a concern of historical linguistics, from its biological, or neo-Darwinian, evolution, a concern of developmental biology.

This distinction was first clearly formulated by Chomsky (1965, 1986), who recognized that de Saussure's definition of language as a property or product of a social group did not lend itself to treatment in biological evolutionary terms. Chomsky made the necessary move by reformulating the langue-parole distinction as competence (what a speaker-listener knows) and performance (what a speaker-listener does in implementing knowledge). He thus set the locus of human language capacity in the individual's
mind/brain. He also set the stage for the modern study of language acquisition.

Form and function

Nonetheless, the competence-performance distinction as usually formulated also does not lend itself to evolutionary treatment. Competence is knowledge of a particular language, formed by interaction between an innate schema, the "universal grammar," and the grammar of the language being learned. Universal grammar is "...a theory of the 'initial state' of the language faculty, prior to any linguistic experience" (Chomsky, 1986, pp. 3-4). Performance is thus said to be a product of a partially innate competence. In what follows, I shall argue that this precisely reverses the true course of development. Universal grammar is not a prescription, or program, for development, but a partial and a posteriori description of the phenotypic product of the developmental system (cf. Oyama, 1985). Universal grammar is a consequence, not a condition of development.

What is at stake here is the relation between form (competence) and function (performance). Zoologists have traditionally stressed the harmonious match between form and function, as expressed in an animal's mode of life, but they have disagreed on how the match comes about. From classical times down into the 19th century, the standard belief was that species and genera had fixed, unchanging forms or essences: departures from the species prototype were "unreal," and structure took precedence over function. For example, according to Mayr (1982), Georges Cuvier (1769-1832), the great French zoologist, held that "...structure has primacy over function and habit, and ...only a change in structure might necessitate a change in function" (p. 367). Certain non-Darwinian French zoologists still hold such views: "...evolution originates in parent forms; if these are absent, new types of organization never appear" (Grassé, 1977, p. 75). Thus, Chomsky and his structuralist forebears in linguistics align themselves with the essentialist tradition in biology by asserting the primacy of form over function.

For Darwin and modern evolutionary biologists, working within the British empiricist tradition, the form-function relation is reversed. Species are not eternally fixed. A species is a genetically variable population of individuals, adapted to a particular ecological niche and thereby reproducively isolated. No aspect of an animal's structure determines a unique function. Rather, a structure determines an unbounded range of functions, to some of which it is more nicely tailored than others. (We can use a screwdriver to drive screws, or as a dagger, a lever, a drumstick, a fork...). If certain members of a species are forced by competition with their fellows, or by an environmental change, into a new mode of life, a new habitat, they may call on some hitherto unused potential function of their structure. The new mode of life then confers a reproductive advantage on those individuals whose structure is marginally better suited (due to small differences in their genetic make-up) to the new mode. Little by little, as with the beaks of Darwin's famous finches (Lack, 1947), the new selection pressures reshape the old structure to its new function. Behavior is thus the great "pacemaker of evolutionary change" (Mayr, 1982, p. 612), the cause not the consequence of speciation and of species form.

In short, a commitment to gradual evolution by natural selection entails a commitment to the primacy of function over form. My central assumption, then, is that language competence (the neural substrate of language form) was shaped in phylogeny, and is still shaped in ontogeny, by language performance (function, behavior).

Behavior as the pacemaker of development

If we extend Mayr's dictum on evolution to development, we see development in a new light. We are freed from preoccupation with the "initial state" as an index of genetic endowment, and from the nativist-empiricist controversy that has dominated studies of language development almost since their inception.

At birth, an organism suffers an abrupt change in the quality and quantity of environmental conditions that may affect its growth. But the discontinuity does not change the developmental system. Gene action does not cease at birth, nor does regulation of gene action by the cellular environment. Rather, the developmental circle widens. The new external environment affords a broader range of stimulation, eliciting a broader range of response. Changes in the organism's environment and behavior now mediate, to a steadily increasing degree, changes in the cellular environment that controls genetic action.

The developmental course is not one of simple maturation (cf. Borer & Wexler, 1987). Language development will not go forward merely because the environment meets general conditions of
survival—air, food, other people, and so on. Specifically linguistic input and, probably, output are called for so that if either is set to zero, development stops. Specific extralinguistic processes of perceptuomotor and cognitive development—a growing grasp on the physical and social modes of being and acting that language represents—must also feed into the language system. Thus, the proper study of language growth is the sequence of behavioral and cognitive conditions, linguistic and extralinguistic, that precipitate language change. The task of this “post-natal embryology,” as we might call it, is to chart the course by which perceptual and motoric functions induce structure, from undifferentiated infant performance to differentiated adult competence.

In short, I am proposing that principles of growth, generally accepted in the development of the peripheral anatomy, also apply to the central nervous system. For example, writing of the plasticity of growing bones, D'Arcy Thompson (1917/1961) remarks “...the very important physiological truth that a condition of stress, the result of a strain, is a direct stimulus to growth itself. This indeed is no less than one of the cardinal facts of theoretical biology. The soles of our boots wear thin, but the soles of our feet grow thick the more we walk upon them: for it would seem that the living cells are 'stimulated' by pressure, or by what we call 'exercise,' to increase and multiply” (p. 238, italics in the original).

Bosma (1975) invokes this principle of exercise in describing the development of the vocal apparatus. Perhaps the development of the neural representations of speech and language is also an instance of “...one of the cardinal facts of theoretical biology.”

Ontogeny (sometimes) parallels phylogeny

Complex functions, and the physical structures that support them, arise in evolution by gradual differentiation from simpler forms. Every evolutionary change is a change in development that is preserved in later generations. Accordingly, when a complex function is an evolutionarily coherent, hierarchically developed system, ontogeny may parallel, or recapitulate, phylogeny. I should emphasize that I am not proposing to reinstate the discredited “biogenetic law” of Ernst Haeckel (1834-1919). I am merely drawing attention to developmental facts well-known since the embryological studies of Karl Ernst von Baer (1792-1876) in the early 19th century (Gould, 1977).

Of course, if we lack, as we largely do for language, precursor forms that confirm the sequence, we cannot be sure that the order of development we now observe was the actual order of evolution. Such precursors as we may establish—for example, lateralized systems for neural control of body posture and manual function (MacNeilage, in press), or certain capacities for symbolic representation (Savage-Rumbaugh, in press)—do not help us in the present context. Some “stages” may have been inserted into the evolutionary sequence later than others that now follow them in development. This may be true, for example, of left hemisphere sensitivity to speech at, or soon after, birth because it is unlikely that a specialized neural substrate for speech evolved before speech itself.

Yet other processes may have evolved independently, parallel to processes with which they were later integrated. For example, some form of segmental phonology, affording at least a modest lexicon, would seem necessarily to have evolved before syntax began to take shape, and we still observe this sequence in development. But later stages of phonology and syntax may have evolved, as they still seem to develop, more or less independently. Similarly, prosodic variations in pitch, amplitude and duration, characteristic of both human and non-human systems of communication, perhaps first followed an independent course of evolution, to be modified and integrated into the linguistic system only as longer utterances and more finely differentiated syntactic functions emerged. The double dissociation of right and left hemispheres for emotional and linguistic uses of prosody may reflect such a course of evolution.

Yet further limits on a phylogenetic interpretation of language ontogeny may seem to arise because the child is born into a community of companions who already speak a language. Moreover, the child is born with a vocal tract and a pair of hands that soon mature into forms adaptable for spoken or signed language, and with a rich, plastic neural substrate fit for shaping by cognitive and linguistic function. The exogenous and endogenous conditions of development therefore differ radically from those that must have prevailed in a hominid community where language was not yet fully formed.

Nonetheless, language acquisition is not instantaneous. The child does not have immediate access to the full adult language that surrounds it.
The effective linguistic environment changes, step by step, as the child comes into possession of new cognitive and linguistic capacities. The problem of how linguistic input is ordered so as to ensure coherent development is solved by the child's own increasingly differentiated linguistic attention. With each new step the child finds the next step waiting, as it were, in the adult language because the adult language is adapted to the child no less than the child to the language. The reason for this is simply that language, like every evolved form, is the product of successive ontogenies, its structure a record of its own evolution (cf. Locke, 1983). Looked at in this way, language is not an object, or even a skill, that lies outside the child and has somehow to be acquired or internalized. Rather, it is a mode of action into which the child grows because the mode is implicit in the human developmental system.

We conclude that language, as a complex, hierarchical, behavioral structure with a lengthy course of development, is a good candidate for (circumspect) study in a recapitulatory framework because its development is rich in sequential dependencies: syllables and formulaic phrases before phonemes and features (as I shall argue below), holophrases before words, words before simple sentences, simple sentences before lexical categories, lexical categories before complex sentences, and so on. Thus, if we assume that each of the subsystems, phonology and syntax, evolved hierarchically by repeated cycles of differentiation and integration, we may recover their course of evolution by tracing the course of their growth. The general heuristic value of the assumption is that it not only throws light on evolution, but also promises an understanding of development in functional terms.

**Development is not teleological**

If we combine the principles of gradual evolution and of a limited, functional recapitulation, we are freed from the temptation to assign purpose to development. At each point in its development an organism is already complete, adapted and adapting, as best it can, to present conditions, internal and external. Just as earlier evolutionary forms existed for themselves, not for any later forms to which they might give rise, so the present form of a developing organism has its own present function. A child does not learn its first words so that it may later combine them into sentences. First words have their own economy. More generally, a child's grammar, at each stage of development, is a possible adult grammar.

This last statement is limited exactly to the extent that the principle of recapitulation is limited. Some stages may have been inserted into the developmental sequence relatively late in evolution. Others may now serve a new function, having lost the function for which they originally arose. Such stages, perhaps particularly during the early months of rapid growth before linguistic functions have begun to differentiate, would have survived through successive generations because they facilitated later adaptive changes, while not infringing on present functions. In this sense, an early form may "preadapt" for a later. However, it would be an error to see such preadaptations as literal preformations. Development, like evolution, is a tinker, putting to present use whatever chance has laid to hand (cf. Jacob, 1977).

**Individuals reach the same developmental ends by different routes and at different rates**

Many linguists believe that universal aspects of language are "innate." Universals are said to be purely genetically determined, and either to be present at birth (Chomsky, 1986) or to mature without contribution from the environment (Borer & Wexler, 1987). Aspects that vary across languages, though perhaps constrained to a few values by principles of universal grammar, are said to be learned from, or "fixed" by, the surrounding language. This is the model of "parameter setting" in syntactic theory (Chomsky, 1981; Roeper & Williams, 1987) and, in effect, of Jakobson's (1968) account of phonological development (cf. Goad & Ingram, 1987).

I do not challenge the descriptive adequacy of such theories. What I question is the proposed developmental mechanism, the attribution of species invariance to the genes, within-species variability to the environment. For, in fact, both invariance and variability arise from both genes and environment. The point is important because individual similarities in language development are often taken as support for a "biological" account of the process, while individual differences are seen as a threat (Goad & Ingram, 1987).

Consider, first, that identical phenotypes may be induced either by a genetic change in an unchanged environment or by a changed environment acting on an unchanged genome. For example, a yellow (as opposed to grey) body in the fruit fly (Drosophila melanogaster) can be produced either by mutation or by feeding the fly larva on fruit jelly impregnated with yellow dye (Dobzhansky, 1957). Similarly, the lack of a
posterior cross-vein in fruit fly wings can result either from mutation or from subjecting the fly pupa to a brief high-temperature shock early in development (Waddington, 1975, Chapters 7 and 8). Another example, not due to experimental intervention, comes from Piaget (1978) who studied fresh water snails in Swiss lakes during the late 1920's. He describes three species. One species (*Limnaea stagnalis*), found in deep, calm waters, has an elongated shell; two other species (*L. lacustris* and *L. bodamica*), found in shallow, rough waters, have almost indistinguishable stubby, contracted shells. If the two stubby species are bred in the calm water of a laboratory aquarium, *L. bodamica* retains its stubby shell over many generations, while *L. lacustris* gradually takes on the elongated form characteristic of *L. stagnalis*. Had Piaget simply described the two stubby species, he might have attributed their similarity to shared genes. But we cannot reliably infer genotype from phenotype, as discussions of language development often assume. Without controlled breeding studies, we cannot determine whether an invariant, species-typical form reflects genetic constancy or environmental constancy.

Notice, moreover, that even when we have isolated genetic or environmental factors that contribute to some aspect of phenotypic variation, we have not demonstrated that the variable held constant has no effect. The snail species that takes an elongated form after generations in calm water, a contracted form after generations in rough water, can adapt to the environmental change because it is genetically equipped to do so. The snail species that retains its contracted form, whether bred in rough waters or in calm, lacks this genetic potential. But we cannot conclude from this that the environment has no effect on the form of its shell. We can conclude only that the two environments to which the snail was exposed did not suffice to select for a change in form although other environments might have done so.

In short, phenotypic form is both genetically and environmentally determined, a fact of some importance to language development. If we cannot assign phenotypic invariance, even in these "simple" examples, either to genes alone or to environment alone, we are surely not justified in doing so for language universals, or for the within-language constancies in linguistic development that are often cited as evidence for an "innate" language competence.

By the same token, variability in development among children learning the same language cannot be assigned solely to the environment. Certainly, different children, even within the same social class, must be exposed to widely varying patterns and frequencies of linguistic input, and such differences are likely to affect the course and rate of development. At the same time, children differ genetically, and these differences too affect development (Locke & Mather, in press). The problem is not to assign variability to a genetic or environmental source, but to understand how children resist the effects of genetic and environmental variation so as to arrive at a common language.

**Development is buffered against extreme variation**

The problem of uniformity within and across languages is, in some respects, the obverse of the problem of language diversity. Languages, like species, differ because individuals differ. If all members of a species were genetically identical, followed an identical course of development, and arrived at an identical developmental term, new species could never arise. Genetic variation within a species, affording subtle individual differences in adaptive response to new environments, is the basis for the origin of species by natural selection. Similarly, if all speakers of a language followed an identical course of development to an identical linguistic term, new languages could never arise.

Of course, differences among speakers that lead to language change are unlikely to rest on genetic differences, because speakers of different languages do not differ systematically in their language-related genetic endowment: children learn any first language to which they are exposed. In fact, from a genetic point of view, language differences rest on commonalities rather than differences: the shared capacity to adapt to any linguistic environment and to learn the surrounding language. In other words, it is precisely because language is supported by (we must presume) many thousands of genes which depend on the environment to trigger their expression, that languages can be learned—and that languages differ. Subtle, environmentally induced individual differences in language development, selected and transmitted to other individuals by sociocultural forces of which we understand at present, very little, are one source of language diversity.

What then are the forces that resist change, guiding speakers into a common language? The problem is a special case, analogous to the general problem in developmental biology of how species
maintain their identities across generations. We can gain conceptual leverage on the question by invoking the principle of "canalization" (Waddington, 1957, 1975). The term is purely descriptive, intended to capture "...a large number of well-known facts in genetics and embryology, all of which are summarized in the statement that the development of any particular phenotypic character is to some extent modificable, and to some extent resistant to modification, by changes either in the genotype or in the environment" (Waddington, 1975, p. 72). In other words, both genetic and environmental variations are buffered against extreme divergence, canalized toward equifinality by constraints from the developing system. Genes and environment act in reciprocal, mutually constraining concert to assure a stable trajectory of growth.

To characterize the constraints on language growth in any detail is well beyond the scope of present knowledge. There would seem, however, to be two main sources of constraint, one in the child itself, one in the physical, social and linguistic environment. Within the child, each step in cognitive and linguistic development opens new paths and closes others, an automatic consequence of the selection, and increasing differentiation, of the neural structures that support language and cognition, continuing the processes of tissue differentiation in the embryo. Once launched, the learning of a particular language becomes increasingly constrained by structural changes in the child's brain. These changes open paths into the language being learned, and perhaps into other languages of the same general type. Presumably, development along these paths will cease if the child is abruptly transferred to a markedly different linguistic environment. (Studies of "savings," that is, of the head-start or lack of it, displayed by children who have switched languages or been exposed to a second language, might then throw light on cognitive and linguistic commonalities between the languages and on their inferred neural substrates.)

The external constraints on the child arise, first, from a general social context that invites the child to engage with its companions and to match its behavior to theirs. The second source of constraint is the language that the child's companions speak. Every language is a solution, one of an uncountable, but presumably limited, set of solutions to the problem of developing a communicative system within the perceptuomotor, memorial and cognitive limits of humans (cf. Lindblom, 1983). The child is then easily guided into its language because, as we have already remarked, every language has evolved under constraints very like those that limit the child.

Given the increasing strength of endogenous and exogenous constraints as the child grows, we would expect phenotypic variation in phonology and syntax, if not in lexicon, to diminish as children come into possession of their language. Nonetheless, individual variability, far from being evidence against a biological account of language development, is a hallmark of developing biological systems.

In the following section, I will attempt to illustrate some of the principles sketched above as they might apply to the development of phonological form over the first two years of life.

THE EMERGENCE OF PHONETIC SEGMENTS

Perception

The first systematic studies of infant capacity to perceive speech derived from the well-known studies of adult categorical perception. In a seminal set of experiments, the model for many others, Eimas and his colleagues (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) demonstrated that 1- and 4-month old infants displayed the same pattern of discrimination between syllables on a synthetic voice onset time (VOT) continuum as adults: they discriminated significantly better between syllables belonging to different English phoneme categories than between syllables belonging to the same phoneme category. Much the same result came from later infant studies of stop consonant place of articulation, consonant manner, the [r]-[l] distinction, and almost every other synthetic continuum on which infants were tested (see Aslin, Pisoni, & Jusczyk, 1983, for a comprehensive review).

From this accumulated wealth of evidence Eimas (1975) concluded that categorical perception reflects the operation of innate mechanisms, specialized for processing speech, and that "...these early categories serve as the basis for future phonetic categories" (p. 342). This conclusion is often cited by students of language acquisition whose primary interest is in syntax. For example, Gleitman and Wanner (1982) summarily dismiss the problem of the origin of discrete phonetic segments (a problem no less severe than the problem of the origin of morphemic units to which they devote much
attention) by citing the work of Eimas and others to support the claim that “...no learning apparatus is required for an initial segmentation of the acoustic wave into discrete phones. The segmentation has been provided in the nervous system” (p. 16). Exactly how this capacity for segmentation came to be evolutionarily, or comes to be ontogenetically, “provided in the nervous system,” they do not consider.

In any event, many reasons to doubt Eimas’s interpretation of the infant data have been given by Jusczyk (1981), Kuhl (1987), Studdert-Kennedy (1986), and Walley, Pisoni and Aslin (1981). I will not rehearse the reasons in detail, but here briefly note three. First, we now know that categorical perception is peculiar to neither speech nor audition (see Harnad (1985) for several relevant papers). Second, Kuhl and her colleagues have demonstrated categorical effects on labial, alveolar and velar VOT continua in chinchillas (Kuhl & Miller, 1975), and on a place continuum in macaques (Kuhl & Padden, 1982, 1983). Thus, whatever the bearing of categorical studies on the development of speech perception, they clearly do not reflect a perceptual specialization. Third, the claim that categorical perception reveals the basis for future phonetic categories confuses two types of category. The categories mimicked by a synthetic series vary along a single acoustic dimension in a fixed context; they comprise, at most, the random variations that we might observe in a single syllable, spoken repeatedly with identical stress and at an identical rate by the same speaker. However, the phonological categories that a child must form are equivalence classes of intrinsic and extrinsic allophonic variants, formed by execution of a particular phoneme in a range of phonetic contexts, spoken with varying degrees of stress, at different rates, and by different speakers.

For study of the infant’s potential grasp on these equivalence classes, we must turn to another body of research. In a systematic series of studies Kuhl and her colleagues have demonstrated that 3- to 6-month old infants can learn to recognize the equivalence of: (i) isolated vowels spoken by a male, a female or a child; (ii) syllables with the same consonantal onset before different vowels; (iii) syllables with a particular consonantal acoustic pattern in initial, medial or final position; (iv) syllables that share an initial “feature” (stop, nasal). (For a comprehensive review, see Kuhl, 1987). Yet we should be cautious in interpreting even these studies as evidence that infants can recognize phonetic (as opposed to acoustic) invariants, if only because Klunder, Diehl and Killeen (1987) have successfully trained Japanese quail to form equivalence classes across syllables spoken with different pitch contours, by different speakers or with the same consonant before different vowels. In short, we have no grounds for claiming that perceptual analysis of the acoustic structure of a spoken syllable, and the formation of categories from the resulting components, engage distinctively human capacities, whether infant or adult.

In fact, these infant studies seem to have no more than a general bearing on the specialized development of language. They are psychophysical studies, demonstrating that infants at, or soon after, birth have the capacity to discriminate and categorize certain acoustic patterns which occur in speech. They provide detailed support for a general observation, suggested by the fact that speech sounds are concentrated in the few octaves of the acoustic spectrum to which humans (and many other animals) are most sensitive: spoken language has evolved and develops within the constraints of prelinguistic auditory capacity. Surely, it would be surprising if this were not so. We do not expect an animal to have a communication system that is not matched to its sensory capacities.

However, the most serious objection to the standard interpretation of these studies is that by assuming the child to be innately endowed with sensitivity to the smallest phonetic units of which spoken utterances are composed, they implicitly adopt a view of development as proceeding from the specific to the general rather than the reverse. Yet, as we have already seen, it is a general rule of both phylogeny and ontogeny that complex structures evolve by differentiation of smaller structures from larger. Accordingly, we should not expect children to build words from phonemes as adults do; rather, we should expect phonemes to emerge from words. We may note, in passing, that a similar principle must apply to the development of word classes and syntactic structures, a fact not generally recognized in developmental psycholinguistics.

From this vantage we can take a different view of infant speech perception than has hitherto prevailed, one that emphasizes development of function rather than psychophysical capacity. One aspect of this work will entail charting the process of “attunement” to the surrounding language, the course by which the infant learns to perceive
speech, bringing to bear capacities evidenced in the laboratory on the speech it hears at home. Work along these lines has, in fact, already begun, with studies of infants' loss of sensitivity to phonetic contrasts not deployed in their native language (e.g., Best, McRoberts, & Sithole, 1988; Werker & Tees, 1984). Other studies by Hirsh-Pasek and her colleagues (e.g., Hirsh-Pasek & Golinkoff, in press; Hirsh-Pasek, Nelson, Jusczyk, Cassidy, Druss, & Kennedy, 1987) are tracing the development of sensitivity to prosodic patterns that specify clausal units, perhaps the thin end of an infant wedge into syntactic structure.

Here, however, I wish to pursue another aspect of perceptual function: its role in guiding the development of production. For I assume that the development of speech depends not only on the maturation and use of the vocal motor system, but also on the infant's gradual discovery of structure in the speech it hears: By learning how to listen the infant learns how to speak. The speech signal therefore comes to specify for the infant the actions—or, more exactly, the motoric components of articulatory action—by which speech is produced. Yet neither phonemes nor features, the perceptual units typically posited in infant studies, can be defined either acoustically or motorically. They are abstract units beyond the reach of an infant who does not yet know a language. If we are to understand speech development, we must couch our descriptions and frame our experiments in terms of auditory and motoric units to which an infant might reasonably be expected to have access. In what follows I will briefly sketch a speculative account of the process by which infants gradually harness prelinguistic motoric elements to communicative use.

Production

We may discern at least four broad stages in the early development of speech, successive cycles of differentiation and integration that carry the infant from prelinguistic cries and mouthings to the emergence of phonetic segments and the beginnings of phonology in early words. The stages form a necessary hierarchical sequence from the general to the specific, from the non-linguistic to the linguistic. However, they are not sharply delimited: processes that begin in one stage may continue more or less unchanged into several later stages—for example, a child may continue to babble long after it has begun to produce words. Also, we must be cautious in characterizing a stage as differentiative or integrative for two reasons. First, because differentiation at one level (for example, breaking a syllable into its component gestures) entails integration at another (coordinating movements to form gestures), and vice versa. Second, because even within a level both differentiation and integration go on during a single stage: for example, integrating gestural patterns into consonants and vowels entails simultaneously differentiating a syllable into its segmental components.

Nonetheless, since even a coarse and tentative taxonomy of development may lend insight into its process, let us consider the following four stages: (1) early vocalizing: differentiation of respiratory and vocal tract activities into patterns of soundmaking associated with different nonspeech actions (0-7 months); (2) canonical babbling: integration of nonspeech movement patterns into rhythmic syllabic structures (7-10 months); (3) variegated babbling and early words: differentiation of the syllable into its component gestures (10-15 months); (4) integration of recurrent gestural patterns into canonical phonetic segments (15-24 months). The suggested time periods are, of course, approximate, because children differ widely in the time courses of their development.

Differentiation: Early vocalizing

Over the first half year of life infant sounds progress systematically from clearly nonspeech vocalizations to canonical babbling that invites phonetic transcription. The changing forms are presumably determined, at least in part, by general maturation and by exercise of the vocal apparatus and of its neural control structures (for review, see Kent, 1981). Stark (1986) divides the development of soundmaking over the first 7-8 months of life into three stages: reflexive crying and vegetative sounds (0-8 weeks), cooing and laughter (8 to 20 weeks), and vocal play (16-30 weeks). These stages are of interest in the present context because they reflect changes in the topography of vocal tract activities and in the forms of the resulting sounds by which prelinguistic elements are marshalled for protolinguistic use (cf. Oller, 1986).

Reflexive crying, like the distress calls of other animals, has, of course, a communicative (though non-linguistic) function. Cries, executed with a relatively unconstricted vocal tract, are predominantly voiced, often with the formant structure of low to mid front vowels. Vegetative
sounds—the grunts, sighs, clicks, stops and pops, associated with breathing and feeding—may be either voiced or voiceless, formed with either an open, vowel-like or a constricted, consonant-like configuration of the vocal tract. The cooing, or comfort, sounds of Stark’s second stage tend to occur in a series of 3-10 segments, each of about 500 ms, separated by voiceless intakes of breath and glottal stops. Their energy is concentrated in frequencies below about 1500 Hz, a pattern that Oller (1980) has termed a “quasi-resonant” nucleus, having the form of a nasal consonant or nasalized vowel.

Thus by the end of the fifth month, the infant’s sound repertoire already contains a variety protoconsonantal and protovocalic elements. The important changes in the next stage are not so much in the size and quality of the repertoire as in its function and organization. First, sounds begin to lose their original functional moorings: they become “...divorced from their previous cry, vegetative or comfort sound contexts, and are used in a variety of communicative situations” (Stark, 1986, p. 159). Second, sounds become longer (700-1500 ms), and form longer, more complex sequences. The infant emits repetitious strings of consonant-like clicks, trills, friction noises, syllabic nasals with a constriction at the front of the mouth, and lip smackings. Vowel-like sounds, now with a “fully resonant” nucleus (Oller, 1980) often carrying extreme pitch glides, are executed with increased variation in tongue height and front-back placement.

However, the key change during this stage is combinatorial: the infant begins to superimpose movements of tongue, jaw and lips on the laryngeal actions associated with cry (Koopmans-Van Beinum & Van der Stelt, 1986), so that the proportion of supraglottal to glottal articulations gradually increases (Holmgren, Lindblom, Aurelius, Jalling, & Zetterström, 1986.) Even at this early stage, children may differ quite sharply in their preferred types of vocalization (Stark, 1986). Toward the end of the stage increasingly long and complex combinations of tract constrictions and openings appear, forming sequences that Oller (1980) terms marginal babble.

The functional value of the differentiation of early soundmaking into these diverse patterns is not obvious. Teleologically, of course, exercise of the vocal apparatus must contribute to its neural and anatomical development, laying the basis for later integration of sounds into syllables. However, the immediate function may simply lie in the increased range of emotional expression that it affords, with a consequent tightening of dyadic social bonds (cf. Stern, 1985; Trevarthen & Marwick, 1986). Interestingly, MacNeilage (personal communication) has noted a possible precursor of such differentiation in the repetitive girneys (lip smacks, accompanied by a low murmur) exhibited by Japanese macaques in intimate, affiliative situations (Green, 1975).

In any event, I have dwelt on this early stage because it is here that we can see most clearly the tinkering together of disparate non-linguistic patterns of vocal tract activity into the beginnings of phonological structure (cf. Bates et al., in press).

Integration: Canonical babble

With the onset of canonical babble, often a sudden event over a few days in the 7th or 8th month, the infant begins to integrate patterns of vocal tract constriction and opening into unitary, cohesive syllables. Rhythmic, reduplicated sound sequences are common: [bababa], [nenene], [diddid]. However, the convenient use of phonetic transcription should not mislead us into supposing that the infant has independent control over segments within a syllable. In fact, rhythmic lowering and raising of the jaw in canonical babble seems to occur with little or no independent movement of the tongue (Davis & MacNeilage, 1990). This rhythmic jaw oscillation often begins about the same time as rhythmic movements of the legs and arms (Thelen, 1981; in press), and perhaps facilitates the integration of vocal tract activities into cohesive syllabic patterns.

We owe the clear distinction between early vocalization and canonical babble to Oller (1980, 1986). He describes some half dozen acoustic properties of structures that listeners recognize as canonical, adult consonant-vowel (CV) (constricted-to-open) or CVC syllables. Of these properties the most important in the present context (because they are the only ones that appear rarely, if ever, in early vocalizations) are temporal. The canonical syllable has a duration of no less than 100, no more than 500 ms; onset (and, when present, offset) formant transitions display smooth changes in frequency and amplitude with durations between 25 and 120 ms. These acoustic patterns reflect the infant’s increasing skill in integrating the closing and opening phases of jaw movements.

Patterns of movement in babble are not random. During the constriction phase of the syllable, complete closure, as in stops, is favored over
partial closure, as in fricatives. Points of closure are biased toward the front of the mouth, engaging lip and tongue tip muscles active in sucking. During the open phase, the favored tongue positions are those associated with low front vowels, indicating that the tongue tends to ride up and down on the jaw with little or no active movement of its own (Davis & MacNeilage, 1990). The glottis is typically approximated throughout the syllable, giving the impression of voiced consonants at syllable onset.

These “phonetic” biases have been reported for infants growing in a number of language environments (Locke, 1983). The biases are also present in many adult languages, perhaps reflecting the infant proclivities from which languages have evolved (Locke, 1983; cf. Lindblom, 1989). Here we have a tangle that cross-linguistic data on babble are still too sparse to resolve. How far do the perhaps universal biases of infant babble reflect the natu­rational state of the infant vocal apparatus, and how far do they reflect the surrounding language? We may ask much the same question concerning individual differences in babbling repertoire within a language: do they reflect differences in the development of the vocal apparatus or differences in the language patterns which the infants happen to have heard? Whatever the answers to these questions, recent work with deaf infants has shown that the emergence of canonical babble is not purely an effect of maturation.

Deaf infants were once said to exhibit the same babbling patterns as hearing infants, at least over the first year of life (Mavilya, 1969; cf. Locke, 1983). However, recent comparisons have demonstrated that canonical babble does not appear on schedule in deaf infants (Oller & Eilers, 1988; Oller, Eilers, Bull, & Carney, 1985; Stoel-Gammon & Otomo, 1986). Whether this is due to the lack of auditory input from the infant’s own vocalizations, from those of its adult companions, or both, we do not know. If self-stimulation were the only essential, a purely maturational account could still hold. To the extent that communicative interchange, or the impulse to imitate the actions of conspecifics plays a role, the onset of canonical babble would be determined, at least in part, by experience of a surrounding language. Persuasive evidence for the role of the surrounding language comes from reports that deaf children exposed to sign language begin to “babble” with their fingers at about the same age as hearing infants begin to babble with their mouths (Newport & Meier, 1985; Laura Petitto, personal communication).

Finally, what is the function of canonical babble? “None” has been the answer of some (e.g., Jakobson, 1968; Lenneberg, 1967) who viewed babble as random mouthing, sharply discontinuous from truly linguistic utterance. However, Locke and Pearson (1988) have recently reported a severe (though not irremediable) delay in the development of speech in a tracheostomized child, deprived of the opportunity to babble from 5 to 20 months. This suggests that babble is a necessary step in normal linguistic development. Moreover, several studies have now shown that babble merges smoothly into early words and that the phonetic structure of a child’s first words tends to reflect its babbling preferences (Locke, 1983; Oller, Wieman, Doyle, & Ross, 1975; Vihman, Macken, Miller, Simmons, & Miller, 1985; Vihman & Miller, 1988). The immediate function would seem then to be to continue the “imitative” process of aligning the infant’s communicative skills more closely with those of its adult companions.

**Differentiation: Variegated babble and early words**

Toward the end of the first year full syllable reduplication fades. The infant begins to differentiate the closing and opening gestures of successive syllables so that “variegated” sequences (Oller, 1980) appear, giving the impression of variations in consonant and/or vowel (e.g., /nen/, /man/, /ded/). Also, at about this time the child produces its first recognizable attempts at adult words or phrases (All gone, What’s that?), usually (in English children) single syllables or disyllables. The two modes of output then proceed concurrently, often for many months, with words gradually coming to predominate. Over this period the child is gradually forging links between its perceptual and productive capacities.

The perceptual ground for this development seems to be laid by the child’s growing attention to words in the surrounding language. Data on early comprehension are scarce, but children evidently accumulate sizeable receptive lexicons before attempting their first words. Benedict (1979) has reported a longitudinal study of comprehension and production in eight children from the age of 9 or 10 months to a point where they had achieved a productive lexicon of 50 words (between 15 and 22 months). In every child comprehension comfortably outstripped production. On the average, the children understood more than 60 words by the time they could produce 10, with a range from 30 to 182 words; and on the average, their receptive lexicon over the period of the study was three
times their productive lexicon (Benedict, 1979, Table I).

The gap between perception and production demonstrates that a perceptuomotor link is not "innate," as some have proposed (Liberman & Mattingly, 1985), but must develop. Many studies have shown that the phonetic forms of early words are similar to those of concurrent babble (see Vihman, in press, for review). A child selects for imitation words that match "vocal motor schemes" (McCune & Vihman, 1987) already present in its babble and avoids words that do not match (Menn, 1983). Thus, the child initially grows into a lexicon, as it were, by discovering correspondences between adult words and auditory feedback from its own babbled output. Because the structure of auditory feedback must correspond, in some fashion, to the structure of motor controls that produced it, the child's recognition of babble-to-word correspondences presumably facilitates growth of perceptuomotor links, and their gradual extension to new words.

Each word (or formulaic phrase) seems to be a prosodic unit (Macken, 1979), its production planned as a whole (Menn, 1983). Evidence for this comes from gestural interactions. A child often fails to execute two different places or manners of articulation within the same word, thus maintaining some of the reduplicative tendencies of canonical babble: dog [gog], lady [jeiji], duck [tat]. Closing and opening gestures also interact. For example, Davis and MacNeilage (1990) report an extensive study of a child's concurrent babbling and speech over the period from 14 to 20 months of age. Their data are replete with instances not only of consonant, but of vowel and even consonant-vowel assimilation. The latter is revealed by the child's preference for high front vowels following alveolar closures and for low, front-central vowels following labial closures. At the same time, these authors also report an inverse relation between consonant and vowel reduplication: where the child succeeds in combating assimilation in the open phases of a disyllable, she often fails to do so in the closing phases, and vice versa. This demonstrates an incipient segregation of consonants and vowels into phonetic classes.

The study by Davis and MacNeilage is particularly important because the child deployed an unusually large lexicon, growing from about 25 words at 14 months to over 750 words at 20 months. Evidently, a child may have a substantial lexicon long before it has fully mastered segmental structure. The principal phonological achievement of this period, then, is internal modification of the integrated syllable by differentiation of its gestural components.

**Integration: From gestures to phonemes**

The final step in the path from mouth sounds to segments is the integration of gestural patterns of syllabic constriction and opening into the coherent perceptuomotor structures we know as consonants and vowels (Studdert-Kennedy, 1987). As is well known, the status of the segment is problematic. For, on the one hand, we can neither specify the invariant articulatory-acoustic properties shared by all instances of a particular consonant or vowel, nor isolate any given segment as a discrete articulatory-acoustic entity within a syllable. On the other hand, across-word metathetic errors in adult speaking ("Spoonerisms") attest to the functional role of segments in the planning and execution of an utterance (Shattuck-Hufnagel, 1983). Moreover, such errors typically entail exchanges between consonants and vowels that occupy corresponding slots in their respective syllables. Since the exchanging elements may be physically quite disparate, it is evident that their exchange is premised on shared function (onset, nucleus, coda) in the formation of a syllable.

There are therefore two aspects to the emergence of segments as elements of word formation in a child's lexicon. First is the grouping of all instances of a particular gesture-sound pattern into a single class presumably on the basis of their perceptuomotor, or phonetic, similarity (e.g., grouping the initial or final patterns of dad, dog, bed, etc. into the class /d/). Second is the distributional analysis and grouping of these gesture-sound patterns into higher-order classes (consonants, vowels) on the basis of their syllabic functions. These processes of phonetic category formation are perhaps analogous to those proposed by Maratsos and Chalkley (1980) for the formation of syntactic categories.

Evidence from across-word metathetic errors for the formation of these classes in young children is sparse. The only systematic data known to me come from Jeri Jaeger (personal communication). She reports her daughter's first across-word metathesis as occurring in her 27th month: ummy takes for tummy aches. This was followed in her 30th month by fritty pace for pretty face, sea tet for tea set, and Bernie and Ert for Ernie and Bert. Jaeger did not report data on the size of her daughter's lexicon at this time, nor on the complexity of her multi-word utterances. But the collection of errors suggests that both were well advanced.
Two possible selection pressures may precipitate formation of consonant and vowel classes. One pressure is toward economy of storage as the lexicon increases in size. We have seen that a child may accumulate an appreciable lexicon of some 750 words without showing signs of independent segmental control (Davis & MacNeilage, in press). But this lexicon is roughly one hundredth of the size that it will eventually become; and it seems reasonable to suppose that, as the lexicon increases, words should organize themselves on the basis of their shared gestural and sound properties. Recurrent patterns of labryngeal and supralaryngeal gesture would thus form themselves into unitary classes of potential utility for recognition and activation of lexical items (cf. Lindblom, 1988; Lindblom, MacNeilage, & Studdert-Kennedy, 1983).

A second possible selection pressure is toward rapid lexical access in the formation of multi-word utterances. Several authors (e.g., Branigan, 1979; Donahue, 1986) have argued that the form of early multi-word combinations may be constrained by the child's limited ability to organize and execute the required articulatory sequences. One such constraint might be a child's inability to produce two successive words with different initial places of articulation. Donahue (1986), for example, describes two strategies adopted by her son in his first two-word utterances. One strategy was to attempt only those words that conformed to his preexisting rule of labial harmony: big book, big bird, big ball were all attempted, but big dog and big cooky were “adamantly refused” (p.215). The second strategy was to circumvent the consonant harmony rule by adopting vocalic words lacking consonants (e.g., where [eə], want [wæ]) as pivots that could be comfortably combined with many of the words already in his vocabulary. Such findings imply that the integration of gestures into independent phonemic control structures, or articulatory routines (Menn, 1983), may serve to insulate them from articulatory competition with incompatible gestures and so facilitate their rapid, successive activation in multi-word utterances.

CONCLUSION

I have proposed that the study of language development might be fruitfully cast in an evolutionary and recapitulatory framework, as a sort of post-natal embryology. Language universals (like other species characters) are the endpoint of development (and evolution), a consequence, not a condition, of learning a natural language. Language function thus determines language form, and form is viewed as an a posteriori description of, rather than an a priori prescription for, development. We are thus freed from the habit of viewing universals as innate, language-specific properties as learned. Both sets of characters are the product of a species-specific developmental system, in which genetic and environmental conditions cannot be separated.

I have attempted to illustrate the approach with a sketch of the process by which the universal phonological categories of consonants and vowels might emerge. They are viewed as deriving by successive cycles of differentiation and integration from prior non-linguistic perceptual and motor capacities.

REFERENCES


FOOTNOTES


†Also University of Connecticut and Yale University.