

13 The particulate origins of language generativity:
from syllable to gesture

MICHAEL STUDDERT-KENNEDY

1 Introduction

Generativity here refers to two 'creative' aspects of normal language use: unbounded scope of reference and freedom from control by identifiable stimuli (Chomsky 1966: *passim*). These two aspects, though obviously independent, are closely related in their origin (as will be argued below) and in their effects. Together they distinguish language from all other forms of animal communication. So far as we know, the vocal repertoires of other species, including our closest primate relatives, are limited to a few dozen calls, associated with present needs, such as food, sex, predators and various social contingencies. Humans, by contrast, can talk about whatever they choose: past, present, or future, concrete, abstract, or even imaginary objects and events.

Not surprisingly, Maynard Smith & Szathmáry (1995) regard the shift from primate call to human speech as 'the decisive step in the origin of specifically human society' (p. 12). They view the step as the latest of eight major evolutionary transitions in the way information is transmitted between generations. Human speech introduced a new code, a new physical medium of transmission, and a shift from largely genetic to largely cultural inheritance. Drawing the familiar parallel between language and the genetic code, Szathmáry & Maynard Smith (1995) observe: 'Grammar enables a speaker with a finite vocabulary to convey an indefinitely large number of meanings, just as the genetic code enables DNA to specify an indefinitely large number of proteins' (p. 231). But they do not ask how or why the two systems have such extraordinary power.

In fact, until recently, the principle that links them was no better understood than a quarter of a century ago, when Roman Jakobson (1970) wrote: 'One could venture the legitimate question whether the

203 *The particulate origins of language generativity*

isomorphism exhibited by these two different codes, genetic and verbal, results from a mere convergence induced by similar needs, or whether, perhaps, the foundations of the overt linguistic patterns superimposed upon molecular communication have been modeled directly on its structural principles' (p. 440). Here, Jakobson framed the issue as a question, but was evidently inclined to reject functional, evolutionary convergence in favour of structural homology. The answer to the question came, however, from William Abler (1989) who recognized that at least three natural systems – chemical compounding, biological inheritance and human language – share a hierarchical structure, based on particulate units.

2 The particulate principle

Abler (1989) dubbed chemistry, biological inheritance and language 'Humboldt systems', because they all conform to Von Humboldt's characterization of language: they draw on combinatorial mechanisms to make 'infinite use of finite means' (Von Humboldt 1836/1972: 70) by 'a synthetic process . . . [that] creates something . . . not present *per se* in any of the associated constituents' (p. 67). Humboldt systems exploit what Abler called 'the particulate principle of self-diversifying systems'. According to this principle, elements drawn from a finite set (e.g., in spoken language: phonemes, words) are repeatedly permuted and combined to yield larger units (words, sentences) higher in a hierarchy, and more diverse in structure and function than their constituents. The particulate units in chemical compounding include atoms and molecules, in biological inheritance genes and proteins.

Abler's (1989) central insight was that Fisher's (1930) arguments concerning the mechanism of biological inheritance could be extended to chemistry and language. Fisher reasoned that, if the characteristics of parents blended, they would be lost in the average of their offspring, and the characteristics of the offspring would lie between, not outside, those of their parents; variation, critical to the process of natural selection, would then diminish from generation to generation. What we see instead is that variation is conserved, or even increased, across generations, and that characteristics of parents are not lost, but (as Darwin knew and Mendel showed) may reappear in later generations due to crossing. Conserved variation across generations and reappearance of parental

characteristics in later generations demonstrate that biological inheritance rests on a particulate mechanism. The essentially unlimited variation that results from that mechanism provides the stuff on which natural selection works. Hence, the diversity of species.

Similarly, the properties of chemical compounds lie outside, not between, those of the elements from which they are formed: hydrogen burns, oxygen sustains burning, but their combination extinguishes fire. Yet the elements themselves can be recovered from the compound by appropriate analysis. Thus, chemical compounds arose by a particulate mechanism, and the resulting diversity of inorganic and organic compounds was then subject to selection under the chemical conditions of the prebiotic earth (Maynard Smith & Szathmáry 1995). Hence, the diversity of terrestrial non-living matter.

Similarly again, if words were formed by blending portions of the acoustic spectrum, or if sentences were formed by blending words, we would rapidly exhaust the communicative potential of speech. The particulate principle, by contrast, affords a vast range of typological variation: unbounded sets of potential phonetic segments, lexical items and lexical combinations, that are then subject to competing perceptual, motoric and memorial pressures selecting among them for cognitive utility, ease of production and ease of comprehension. Hence, the diversity of languages.

The power of the particulate principle does not stem simply from combinatorics – the world is more than a collection of chemical and genetic anagrams – but from the fact that structures at each higher level of the hierarchy have a broader and entirely different range of function in comparison to those below. We cannot derive the properties of common salt from those of sodium and chlorine, nor of a protein from the gene that controls its formation; in language, we cannot derive the meaning of a word from the phonetic elements that compose it, nor the meaning of a proposition from the lexical meanings of its words without regard to their syntactic grouping.

Finally, although the systems form a chronological sequence from the formation of elements to planetary chemistry to biological inheritance to language, they are not (so far as we can tell) homologs: they do not all descend from a single common ancestral system, nor do the later systems seem to stem from the earlier. Rather, Abler (1989) concludes, the particulate principle is an axiomatic property of the physical world to which any self-diversifying system necessarily conforms. The answer

205 *The particulate origins of language generativity*

to Jakobson's 'legitimate question', then, is that the genetic and verbal codes converged on the only mechanism by which a natural system can 'make infinite use of finite means' '[through] a synthetic process . . . [that] creates something . . . not present *per se* in any of the associated constituents'. Genes and the elements of speech are therefore functional analogs, not structural homologs.

3 The logic of the particulate principle in language

In his discussion of particulate inheritance, Fisher (1930) notes the 'remarkable fact that had any thinker in the middle of the nineteenth century undertaken, as a piece of abstract and theoretical analysis, the task of constructing a particulate theory of inheritance, he would have been led, on the basis of a few very simple assumptions, to produce a system identical with the modern scheme of Mendelian or factorial inheritance' (p. 7). Much the same might be said, *mutatis mutandis*, of a twentieth-century thinker undertaking to construct a particulate theory of language.

Let us assume that the goal is to devise a system for symbolically representing an unbounded set of objects, events and their relations as specified in simple propositions, by means of a finite set of discrete signals. Our imaginary theorist would no doubt immediately see that if he assigned each signal to a different object or event, he would soon exhaust the signal set. It would not then be long before he hit on the idea of repeatedly sampling, permuting and combining the signals, and so took the first step into language: the number of different possible meanings would now be limited only by the number of possible permutations and combinations (and by our theorist's tolerance for homonymy). Notice, however, that the enterprise could only go forward in this fashion if the component signals were bleached, as it were, of meaning. For if the same basic elements are to be repeatedly permuted to construct different minimal units of meaning ('words'), the elements themselves must be meaningless. The resulting loss of direct signal-to-meaning correspondence would be without precedent in animal communication, giving rise to a unique characteristic of human language, namely, a level of structure between signal and message, or, for spoken language, phonology.

Once our theorist had hit upon the combinatorial principle to build an indefinitely large lexicon from a few dozen signal units, he might be

tempted to consider his task done, because language users could now continue indefinitely assigning signals to objects, events and propositions concerning their relations. He would soon see, however, that the system placed a prohibitive tax on creative phonetic invention and memory by obliging its users to construct a new word for every new proposition. Under his charge to achieve unbounded representational scope, our theorist would then surely take the second step into language when he saw, for example, that the six distinct signals for 'X', 'X verbs', 'X verbs Y', 'Y', 'Y verbs', and 'Y verbs X' could be reduced to three by breaking holistic signals with complex referents into new basic components of meaning, and marking each for its thematic role in a proposition by means of sequential order, or some other arbitrary device. Thus, our theorist would arrive at a simple subject-predicate syntax by having recourse to the combinatorial principle a second time, introducing a second level of structure between signal and message.

According to the particulate principle, then, the properties that distinguish language from all other modes of communication are the formal structures interposed between signal and meaning. Animal signals, as best we can judge, directly express a sender's needs or emotions and directly elicit a receiver's response. Language, mediated by phonology and syntax, puts computational distance between signal and message, introducing a cognitive buffer impervious to the immediate environmental demands of stimulus or response. Evidently, a system of communication with the unbounded scope of reference afforded by the particulate principle necessarily dissociates signal from meaning, and thereby sets up a cognitive preadaptation, perhaps, for its evolution into a system free from stimulus control. Thus, the two properties that assure language its unique power, though conceptually independent, have a common origin in the particulate principle.

Here, I should emphasize again an aspect of Abler's thesis that can easily be missed. The particulate principle is a property of the physical universe to which any system that 'makes infinite use of finite means' - any system with the scope and freedom of language, and so of human thought - necessarily conforms (Abler 1989, 1997). Once a system of referential communication began to evolve, its evolutionary terminus, in a hierarchical particulate structure, was as mathematically inevitable as, say, the hexagonal form of a honeycomb cell or the logarithmic spiral of a snail-shell (Thompson 1961). The task for a theory of language evolution is not therefore to derive its hierarchical structure. Rather the tasks

207 *The particulate origins of language generativity*

are, first, to understand the selection pressures which forced that structure into existence by demanding a referential lexicon of increasing scope, and, second, to specify the physical and psychological conditions that have shaped properties of phonology, morphology and syntax common to all languages.

The latter task comes into focus if we compare spoken and signed languages. That these two forms of language share the same overall hierarchical structure is, according to the particulate principle, physically and mathematically inevitable. What we have to understand is how differences between the modalities, between hand and mouth, eye and ear, forced language into different surface manifestations (Bellugi & Studdert-Kennedy 1980).

4 Some preliminaries to an account of language evolution

I will not speculate on the complex social and cognitive infrastructure that must have been in place before pressures toward referential vocal communication could arise and take effect. I merely draw attention to Bickerton's (1990, 1995) and Donald's (1991) excellent discussions of these matters. Here I address, rather, two perhaps more answerable questions: what are the basic particles of speech, and how did they evolve? In what follows, I shall argue: (a) that the basic particles of speech are not, as generally assumed, phonetic segments (consonants and vowels) or their descriptive features, but the gestures that form them; (b) that gestures arose evolutionarily by differentiation of holistic, syllable-like vocalizations, and so of the hominid vocal tract and its component articulators, in response to pressure toward an increased vocabulary (see Lindblom, this volume). I will support the argument with data from early child speech. But, first, some preliminaries.

4.1 *The lexicon, precursor of syntax*

The two steps sketched above, first into naming (phonology), then into discursive statement (syntax), correspond exactly to Bickerton's (1990, 1995) two stages of protolanguage and full language. I am not concerned here with dating these stages, with their relation to hominid brain/body ratio, or with their role in advancing hominid culture, but merely with their logical sequence in the evolution of language according to the particulate principle. Bickerton (1995) recognizes that 'syntax could

not have come into existence until there was a sizable vocabulary whose units could be organized into complex structures' (p. 51). But, curiously, he does not recognize that a sizable vocabulary could not have come into existence until holistic hominid vocalizations had been differentiated into phonetic units that could be organized into words. The present thesis therefore supplements and complements Bickerton's by arguing that precisely because a sizable lexicon is a precondition of syntax, the hominid breakthrough into naming by means of a particulate phonetics was the first step into language.

4.2 *Particulation of the vocal machinery*

I take a single phonated cycle in the closed-open oscillation of the primate mandible to be the first unit of referential meaning, a protosyllable (see MacNeilage, in press). The problem for the early hominid, under pressure for an increasing vocabulary, was then to produce a sufficient number of articulatorily and perceptually distinct protosyllables. The critical step into particulation, and so into the vast lexicons of human languages, was evidently accomplished by internal modulation of the spectral properties of the protosyllable.

Some evidence for modulation of rhesus calls by jaw and lip action comes from recent studies by Hauser and his colleagues (Hauser 1996: 182ff.), but the full range of human vocalization rests on evolutionary changes in the hominid supralaryngeal airway (Lieberman 1984) and in the neuro-musculature of larynx, tongue, velum and lips (Lenneberg 1967). These changes transformed the primate vocal tract from a straight, largely inflexible tube to a bent tube with several more-or-less independently movable parts; particularly important was differentiation of the tongue, affording partially independent action of root, body and blade. The resulting articulatory flexibility permitted the vocal tract to become a variable musical instrument, as it were, the equivalent of as many different instruments with their characteristic resonant properties as there are distinct configurations of the vocal tract, each excited by as many different pitches or patterns of airflow as there are distinct settings of the vocal cords. Here, then, is the particulation of the vocal machinery and its neural controls on which the hierarchy of language was raised.

This assertion challenges the familiar observation that: 'It is not the mere power of articulation that distinguishes man from other animals, for as everyone knows, parrots can talk' (Darwin 1871: 54). Although

209 *The particulate origins of language generativity*

their vocal mechanism is not, so far as I know, fully understood, parrots evidently imitate speech in a continuous, non-particulate fashion, much as they imitate fire sirens and lawnmowers. Certainly, a parrot trained to speak German phrases does not then speak novel English phrases with a German accent (Fernando Nottebohm, personal communication). Nor do we have reports of psittacine spoonerisms.

Also well beyond the reach of talking parrots are rates typical of conversational speech (in English, as many as 120–180 words/minute, or roughly 10–15 phonemes/second). Distribution of the communicative load over independently movable parts was, in fact, critical to the execution of movements fast enough to evade the limits of short-term memory (Lieberman 1984; Studdert-Kennedy & Liberman 1963). Such a rate can be achieved

only if separate parts of the articulatory machinery – muscles of the lips, tongue, velum, etc., – can be separately controlled, and if . . . a change of state for any one of these articulatory entities, taken together with the current state of others, is a change to another phoneme . . . It is this kind of parallel processing that makes it possible to get high-speed performance with low-speed machinery. [Lieberman, Cooper, Shankweiler & Studdert-Kennedy 1967: 446]

Similarly, hand configurations in American sign language or in finger-spelling (where rates of 5 letters/second are standard (Wilcox 1992)) depend on rapid movements distributed across wrist and fingers.

4.3 *Segments, features and gestures*

The functional status of phoneme-sized phonetic segments (consonants and vowels) is attested by the alphabet: the speech of any language can be transcribed in alphabetic symbols by a competent listener/writer, and recovered from the script by a competent reader/speaker. Indeed, without the notational system of the alphabet and our intuitive grasp of its principles, study of language would scarcely be possible at all. Nonetheless segments cannot be the basic particles of speech for at least two reasons. First, consonants and vowels have no status outside language; they are linguistic entities, defined by their phonetic function in the formation of a syllable, and are therefore part of what an evolutionary account, undertaking to derive language from its non-linguistic precursors, must explain. Second, consonants and vowels are not primitive units, but compounds, analogs of the molecule, not the atom; according to the standard structuralist formulation, they are 'bundles' of features (Bloomfield 1933: 79; Jakobson & Halle 1956: 8).

Note, however, that the features themselves have no substance. As is evident from the adjectival terminology of all feature theories (nasal, coronal, continuant, etc.), features are not independent entities (like atoms), but properties or attributes of the segments they describe (Fowler, Rubin, Remez & Turvey 1980). In fact, Jakobson and Halle (1956) write, referring to features: 'Phonemic analysis is a study of *properties*, invariant under certain transformations' (p. 13, my emphasis). On the standard account, then, a segment is defined as the sum of its featural properties. On the account adopted here and elaborated somewhat below, a segment is a recurrent constellation of articulatory gestures and their acoustic consequences, of which features may then serve as classificatory descriptors. We thus move toward a substance-based, explanatory account rather than a purely formal account of the origins of speech, and so of language (Lindblom 1980, 1986; Lindblom, MacNeilage & Studdert-Kennedy 1984).

4.4 *Coevolution of speaking and listening*

Language belongs to a large class of complex, interlocking social behaviours in which two or more members of a species differ, but are mutually adapted. Sex is another obvious example. A third is the distribution of functions – provisioning, building, protecting, reproducing – over individuals in an insect colony. Such adaptively complex social patterns presumably reflect the interlocking action of many genes and are therefore incompatible with a saltationist account of their evolution. If a single mutation had granted some hominid the power of speech, it would have had no one to talk to.

Language differs from the other examples because the complementary behaviours (speaking and listening) are lodged in a single individual, and from other primate vocal communication, because the behaviours are learned. Vocal learning is largely confined to a few species of birds and to humans (Hauser 1996). Evolution of the capacity evidently entailed gradually aligning perception and production to achieve 'parity' between sender and receiver (Liberman 1996: 31; Liberman & Mattingly 1989), so that the sounds of one came to specify for the other the motor components deployed in their production, and vice versa (Studdert-Kennedy 1983). Implicit here is the notion that the underlying units of speech production and speech perception are abstract, formally identical, neural control structures (or processes), and this will be my assumption in what follows.

211 *The particulate origins of language generativity*4.5 *Ontogeny epitomizes phylogeny*

For Darwin the facts of embryology were 'second in importance to none in natural history' (1859/1964, p. 450) for the light they could throw on a species' ancestry: 'the embryonic state of each species . . . partially shows us the structure of . . . less modified ancient progenitors' (p. 449). Today, of course, we know that ontogeny does not recapitulate phylogeny in the fashion posited by Haeckel's 'biogenetic law'. That 'law', with its Lamarckian assumptions, was evidently accepted by Darwin himself (Richards 1992), but fell with the rise of Mendelian genetics (Gould 1977). Since all genes are present from the start of development, evolutionary changes do not have to be tacked on at the end of development (as Haeckel's 'law' required), but can be inserted into the sequence at any point.

Consider here the phenomenon of babbling and the early onset of syllable differentiation, documented by MacNeilage & Davis (1990; Davis & MacNeilage 1995). If the assumption that differentiation of the hominid protosyllable evolved in response to pressure for increased vocabulary is correct, the onset of differentiation before the first words in modern children must be a relatively late evolutionary novelty, selected and inserted into the developmental sequence for whatever facilitatory effect it may have on later processes of differentiation.

Nonetheless, parallels between development and evolution do exist for at least two reasons. First, both evolution and development proceed by successive cycles of differentiation. However diverse the developmental paths of individual organisms, the sequence is necessarily from simple to complex. Second, every evolutionary change is a change in development, preserved and passed on to later generations. Even if the exact evolutionary sequence is lost, development is a summary record of those changes (Gould 1977). The next section draws on this fact to sketch the evolution of particulate speech as epitomized in early development.

5 Learning to (p)articulate

We now have to see how development of the particulate principle in the child (and so, by inference, its evolution in the species) arises from and depends on a particulate vocal machinery. We begin with an account of the gesture as currently defined in the developing theory of articulatory phonology.

5.1 *Articulatory phonology*

The term, *gesture*, is often used intuitively to refer to intentional movements of the speech articulators, but recently has been given a precise (if preliminary) definition in the articulatory phonology of Browman, Goldstein and their colleagues at Haskins Laboratories (e.g. Browman and Goldstein 1992). They have incorporated the gesture as the basic phonetic and phonological unit of articulatory action into the only explicit model of speech production currently available. What follows is a brief sketch of that model.

If we watch, or listen to, someone speaking, we see, or hear, the speaker's mouth repeatedly closing and opening, forming and releasing constrictions. In the framework of articulatory phonology, each such event, each formation and release of a constriction, is an instance of a *gesture* (cf. Bell 1911: 38-39). Constrictions can be formed within the oral, velic or laryngeal articulatory subsystems; within the oral subsystem, they can be formed by the lips, the tongue tip or the tongue body. The function of each gesture, or act of constriction, is to set a value on one or more vocal tract variables that contribute to the shaping of a vocal tract configuration, by which (in conjunction with pulmonic action) the flow of air through the tract is controlled, so as to produce a characteristic pattern of sound.

Figure 13.1 displays the tract variables and the effective articulators of a computational model for the production of speech, at its current stage of development (Browman & Goldstein 1992). The inputs to the model are the parameters of sets of equations of motion for gestures; thus, within the model, a gesture is an abstract description of an articulator movement, or of a coordinated set of articulator movements, that unfolds over time to form and release a certain degree of constriction at a certain location in the tract. Settings of the parameters permit constriction degree to vary across five discrete values, corresponding to stop closure, fricative approximation and three degrees of vowel height; constriction location for oral gestures varies across nine discrete places of articulation, from lips to pharynx.

Note that a gestural description is not simply a change in terminology. Gestures do not correspond one-to-one with either segments or features, and a gesture is not, as is sometimes supposed, the functional form of a feature. For example, if we take glottal approximation for

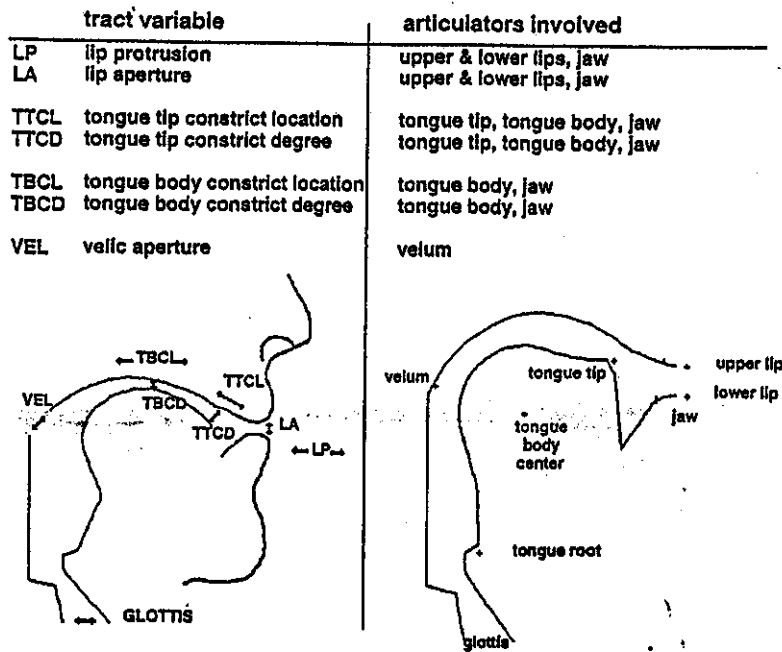


Figure 13.1 Tract variables and articulators in a model of speech production.

voicing to be the unmarked state of the glottis, the initial segment of the word *dip* corresponds to one gesture (alveolar closure), while its nasal counterpart in *nip* corresponds to two (alveolar closure and velic lowering). Of these two gestures, the first corresponds to three features [+consonantal], [+anterior], [+coronal], the second to one, [+nasal].

The gestures for a given utterance are organized into a larger coordinated structure, represented by a gestural score. The score specifies the values of the dynamic parameters for each gesture, and the period over which the gesture is active. Figure 13.2 (centre) schematizes a stripped-down score for the word *nut* ([nat]), as a sequence of partially overlapping gestural activation intervals. Each gesture has an intrinsic duration that varies with rate and stress. Correct execution of an utterance then requires accurate timing of the gesture itself, and accurate phasing of gestures with respect to one another, both sources of 'error' in early child speech.

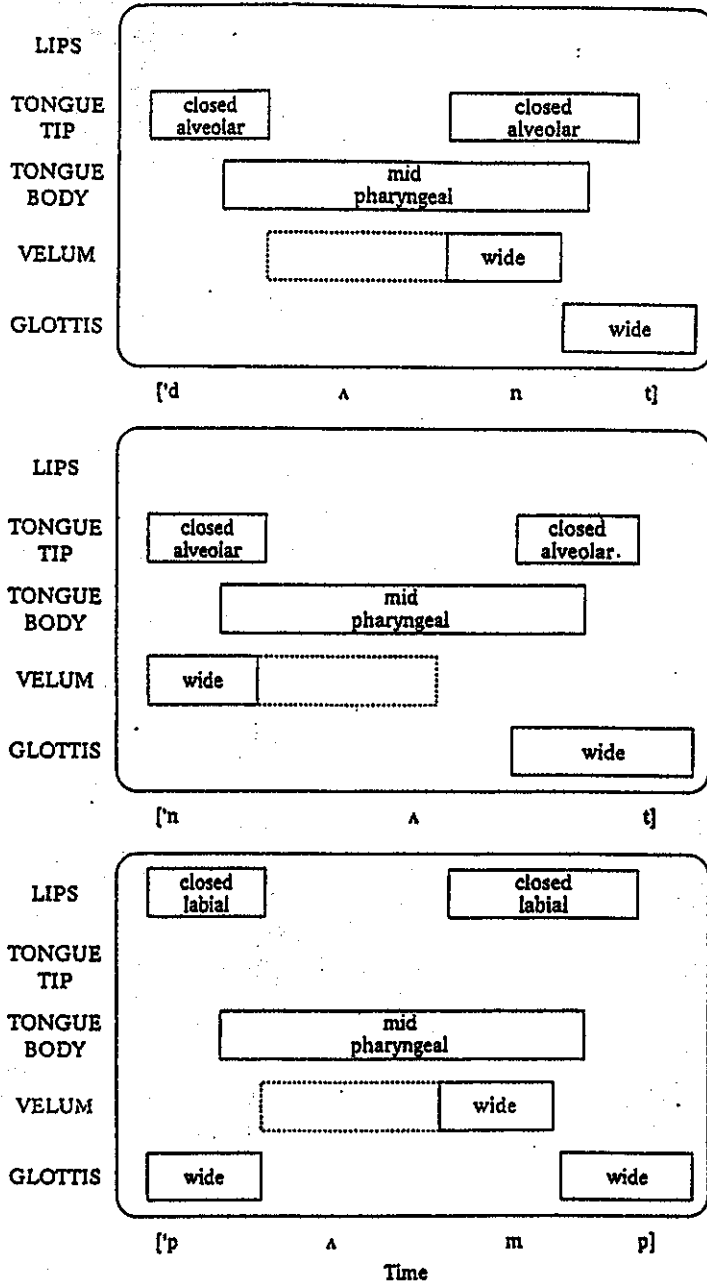


Figure 13.2 Gestural scores for [dant], [nat] and [pamp].

5.2 *Differentiation of the syllable*

The syllable is the central integrative unit of spoken language, the smallest utterance in which segmental phonology, prosody, syntax and meaning can come together. The syllable is also the apparent homolog of certain primate vocalizations (MacNeilage & Davis 1990), and the first unit recognizable as speech in the infant's vocal development. With emergence of the syllable in canonical babble around the seventh month of life, listeners begin to transcribe infant utterances with a fair degree of reliability. We should not infer from this, however, that the babbling infant has independent control over the segmental components of the syllable; the segments are in the ear of the listener, not in the mouth of the speaker.

Indeed, MacNeilage & Davis (1990) have presented evidence and argument that the syllable trains of reduplicative babbling (e.g. [bæbæbæ], [nɛnɛnɛ]) reflect rhythmic up-down movements of the jaw with no internal modulation of the syllable by the tongue or velum. The different clesant (consonant-like) values across utterances (e.g. [b] vs [ɹ] above) reflect differences in the resting positions of tongue and velum at syllable onset, maintained throughout the string; the different vocant (vowel-like) values (e.g. [æ] vs [ɛ] above) reflect different amplitudes of jaw lowering. The importance of this interpretation is that it warrants a basic unit of articulatory action, presumably homologous with the hominid protosyllable, from which the gesture can emerge by differentiation, as the child gains independent control over its articulators.

The first move toward such control comes with the beginning of 'variegated babble' (Oller 1986) in which syllable onsets and/or nuclei vary from one syllable to the next (e.g. ['we:'da'wi:'da'me:'na'mu:'ni]. Yet here the data are ambiguous. Because babbled syllable trains are typically voiced throughout, we usually cannot tell whether the child is executing a variegated string of segments, as MacNeilage & Davis (1990) suppose, or a variegated string of gestures, as the present account supposes. And because the child has no evident adult model, or target, for its babbling, we have no way of inferring the units from its 'errors'. In fact, not until the child's first words can we plausibly infer units of motor control below the level of the syllable.

5.3 *Differentiation of the word*

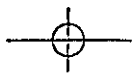
No doubt the vocal play of babble facilitates motor development, but it is communicative intent, or meaning, that starts a child talking. The earliest vocal unit of meaning is probably the prosodic contour; the

earliest segmental unit of meaning (or of segmental sound contrast) is the word (Ferguson & Farwell 1975; Menn 1983; Waterson 1971), and it is by differentiation of the syllables of a word that the gesture first emerges as a phonological unit (Lindblom 1992; Studdert-Kennedy 1987). Evidence for this process comes from the varying phonetic forms with which a child, on the cusp between babble and words, attempts a word slightly beyond its phonetic reach.

At least two facts about these variants argue for the gesture as the basic unit of articulatory action and for the word as the domain over which gestures are organized: (a) variants are often composed of many, or all, of the gestures, but few, or none, of the segments in the target word; (b) erroneous segments in the variants can often be seen to arise from incorrect phasing of correct gestures.

My first example comes from Ferguson & Farwell (1975) who report ten radically different attempts by a fifteen-month-old child to say *pen* [pen] within one half-hour session: [mãə], [ʔ̃], [dè^{da}], [hm], [mbð], [p^hIn], [t^hnt^hnt^hn], [ba^h], [d^hau^N], [buã] (cf. Studdert-Kennedy 1987). On the surface, these variants seem absurdly diverse from one another and from their target. Yet every token includes between two and five of the five gestures in the target (labial closure, glottal opening, tongue tip raising, alveolar closure, velic lowering), although incorrectly phased with respect to one another. For example, lip closure for initial [p^h], correctly executed with open glottis and raised velum, will yield [m^b], as in [m^bð], if velic lowering for [n] is momentarily activated and if glottal closure for [en] is initiated at the same time as lip closure, tens of milliseconds earlier than in the correct utterance. Note that, out of some thirty-seven gestures transcribed in these ten attempts, only ten (27 per cent – those for the vowels [ʌ] [ɪ], [a], [o] and [u]) do not occur in the target; on the other hand, out of the twenty-seven full segments transcribed, twenty-one (78 per cent) do not occur in the target.

My second example, drawn from Studdert-Kennedy & Goodell (1995) describes the attempts of a two-year-old child to say the same syllable, [nat], in the disyllabic words *doughnut* and *peanut*. This syllable elicited quite different patterns in the two contexts and in the same context on different occasions. The pattern of errors demonstrates unequivocally that, while the syllable retains its role as an organizing motor template, or 'frame' (MacNeilage, in press, this volume; MacNeilage & Davis 1990) within the word, the word itself is the domain over which gestures are organized.



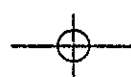
217 *The particulate origins of language generativity*

Table 13.1 Variability within and between words spoken by a two-year-old child: the same target syllable executed differently in different phonetic contexts and on different occasions. The utterances are listed chronologically, but the columns for *doughnut* and *peanut* are not synchronized.

<i>Nut as in doughnut and peanut</i>			
<i>doughnut</i> ['do:nat] --->	<i>nut</i>	<i>nut</i>	<--- <i>peanut</i> ['pi:nat]
['do:'di:dat]	dət	də	['pe:'də]
['du:'dətʃ]	dətʃ	tə	['pe:n'tə]
['du:'dats]	dats	de:	['pe:'de:]
['du:'də]	də	pəmp	['pe:m'pəmp]
['du:n'dant]	dant	pəp	['pi:'pəp]
['do:'nat]	nat	nat	['pi:'nat]

Table 13.1 lists in chronological order some of the variations on *nut* in *doughnut* and *peanut*. Among the child's first attempts at *doughnut* were ['du:'dətʃ] and ['du:'dats]. Word-final alveolar constrictions added apparent fricative segments not present in the model. They were not attempts at the plural, because the child was given only part of a doughnut to eat, only heard the word in the singular, and did not yet command the plural morpheme. Rather, they seem to have resulted from a relatively slow release of [t], making the fricative portion of the release more salient. Subsequent attempts at this word varied over forms as diverse as ['du:'də] and ['du:n'dant]. The consonantal pattern of the latter seems to result from prolongation of the alveolar closure for medial [n] after velic release, giving an unwanted [d], combined with prolongation of the alveolar closure for final [t] and a shift in (or harmonious repetition of) the medial velic gesture, giving the unwanted final cluster. Figure 13.2 (top) displays a schematic gestural score illustrating the errors in alveolar closure duration and in phasing of velic action required to make the shift from ['nat] to ['dant].

For *peanut* the child first tried ['pe:'də], omitting velic action, and a few days later, ['pe:n'tə], where prolongation of the medial alveolar closure, combined with a shift in the phasing of the final glottal opening, relative to velic closure and the tongue body gesture, gives rise to an apparent shift in the ordering of the target consonant-vowel-consonant sequence. Later, she offered ['pi:'pəp], omitting the velic gesture and



succumbing to labial harmony, and ['pe:m'pamp]. The latter, formally analogous to ['du:n'dant] for *doughnut*, with its velic harmony, mistimed velic action and resulting unwanted segments, is further complicated by the substitution of harmonized labial closures for the alveolar closures called for by the target. Figure 13.2 (bottom) illustrates the errors of gestural location and duration and of the phasing of velic action required to make the shift from ['nat] to ['pamp]. My point in these examples is that a featural or segmental account may well describe the form of the errors, but can give no account of the process by which the errors came to be made. By contrast, a gestural account offers a simple description of both the process and the outcome.

6 Summary and conclusion

The generativity, or openness, of language rests on 'the particulate principle of self-diversifying systems', a physical principle, common also to chemistry and genetics, by which discrete particles or elements, drawn from a finite set (in spoken language: phonemes, words), are repeatedly permuted and combined to yield larger units (words, sentences), higher in a hierarchy and more diverse in structure and function than their constituents. Evolution of such a system in primate communication had the unprecedented effect of inserting computational processes between signal and message, a necessary preadaptation for the emergence of language as a system of thought and self-expression, free from identifiable stimulus control. The complex social and cognitive conditions in hominid communities that selected for referential vocal communication are a matter for speculation, but the eventual effect of those pressures was differentiation of the hominid vocal tract to form a particulate vocal machinery, adapted for rapid execution of gestures, the elements of the emerging hierarchy. Because ontogeny is a summary record of phylogeny, and because complex structures necessarily emerge from simpler structures by successive steps of differentiation, we can trace a possible course of hominid evolution by tracking the developing child from the undifferentiated mandibular oscillation of early babble through the imprecise and poorly timed gestural patterns of early words to the rapid articulatory routines of mature speech.

By establishing the gesture as the irreducible element of phonetic perception and action, and by demonstrating that gestures emerge from



219 *The particulate origins of language generativity*

syllables as a child strives to negotiate its first words, we not only offer a model for the origin of the particulate principle in the evolution of speech, and so in language, but also take a step toward grounding language and thought in human anatomy and physiology.

Acknowledgments

My thanks to William Abler for his papers and stimulating correspondence, and to Cathi Best, Noam Chomsky, Louis Goldstein, Marc Hauser, Alvin Liberman and Ignatius Mattingly, for instructive comments on an earlier version. Preparation of this chapter was supported in part by Haskins Laboratories.

References

- Abler, W. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, 12, 1-13.
- (1997). Gene, language, number: the particulate principle in nature. *Evolutionary Theory*, 2, 237-248.
- Bell, A. G. (1911). *The Mechanism of Speech*. New York: Funk & Wagnalls.
- Bellugi, U. & Studdert-Kennedy, M. (eds.) (1980). *Signed and Spoken Language: Biological Constraints on Linguistic Form*. Deerfield Beach, FL: Verlag Chemie.
- Bickerton, D. (1990). *Language and Species*. Chicago and London: University of Chicago Press.
- (1995). *Language and Human Behaviour*. Seattle: University of Washington Press.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.
- Browman, C. P. & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49, 155-180.
- Chomsky, N. (1966). *Cartesian Linguistics*. New York: Harper & Row.
- Darwin, C. (1859/1964). *On the Origin of Species*. (Facsimile edition.) Cambridge, MA: Harvard University Press.
- (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.
- Davis, B. L. & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech & Hearing Research*, 38, 1199-1211.
- Donald, M. (1991). *Origins of Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.
- Ferguson, C. A. & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 51, 419-439.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
- Fowler, C. A., Rubin, P., Remez, R. E. & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In Butterworth, B. (ed.) *Language Production*, vol. 1. London: Academic Press.
- Gould, S. J. (1977). *Ontogeny and Phylogeny*. Cambridge, MA: Belknap Press of Harvard University Press.



220 Michael Studdert-Kennedy

- Hauser, M. (1996). *The Evolution of Communication*. Cambridge, MA: MIT Press.
- Jakobson, R. (1970). *Main Trends of Research in the Social and Human Sciences*. Paris/The Hague: UNESCO/Mouton: 437-440.
- & Halle, M. (1956). *Fundamentals of Language*. The Hague: Mouton.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.
- Lieberman, A. M. (1996). *Speech: A Special Code*. Cambridge, MA: MIT Press.
- Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, 243, 489-494.
- Lieberman, P. (1984). *The Biology and Evolution of Language*. Cambridge, MA: Harvard University Press.
- Lindblom, B. (1980). The goal of phonetics, its unification and application. In *Proceedings of the Ninth International Congress of Phonetic Sciences*, vol. III. Copenhagen: Institute of Phonetics: 3-18.
- (1986). Phonetic universals in vowel systems. In Ohala, J. J. & Jaeger, J. J. (eds.) *Experimental Phonology*. Orlando, FL: Academic Press: 13-44.
- (1992). Phonological units as adaptive emergents of lexical development. In Ferguson, C. A., Menn, L. & Stoel-Gammon, C. (eds.) *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press: 131-163.
- MacNeilage, P. & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. In Butterworth, B., Comrie, B. & Dahl, O. (eds.) *Explanations for Language Universals*. New York: Mouton: 181-203.
- MacNeilage, P. F. (in press). The frame/content theory of evolution of speech production. *Behavioral & Brain Sciences*.
- & Davis, B. L. (1990). Acquisition of speech production: the achievement of segmental independence. In Hardcastle, W. J. & Marchal, A. (eds.) *Speech Production and Speech Modelling*. Dordrecht: Kluwer: 55-68.
- Maynard Smith, J. & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford: W. H. Freeman.
- Menn, L. (1983). Development of articulatory, phonetic and phonological capabilities. In Butterworth, B. (ed.) *Language Production*, vol. II. London: Academic Press: 3-50.
- Oller, D. K. (1986). Metaphonology and infant vocalizations. In Lindblom, B. & Zetterstrom, R. (eds.) *Precursors of Early Speech*. New York: Stockton Press: 21-35.
- Richards, R. J. (1992). *The Meaning of Evolution*. Chicago: University of Chicago Press.
- Studdert-Kennedy, M. (1983). On learning to speak. *Human Neurobiology*, 2, 191-195.
- (1987). The phoneme as a perceptuo-motor structure. In Allport, A., MacKay, D. G., Prinz, W. & Scheerer, E. (eds.) *Language Perception and Production*. London: Academic Press: 67-84.

221 *The particulate origins of language generativity*

- & Goodell, E. W. (1995). Gestures, features and segments in early child speech. In De Gelder, B. & Morais, J. (eds.) *Speech and Reading*. Hove: Erlbaum (UK), Taylor & Francis: 65-88.
- & Liberman, A. M. (1963). Psychological considerations in the design of auditory displays for reading machines. *Proceedings of the International Congress on Technology and Blindness*, vol. I. New York: American Foundation for the Blind: 289-304.
- Szathmáry, E. & Maynard Smith, J. (1995). The major evolutionary transitions. *Nature*, 374, 227-232.
- Thompson, D. W. (1961). *On Growth and Form*. (Abridged edition edited by J. T. Bonner.) Cambridge: Cambridge University Press.
- Von Humboldt, W. (1836/1972). *Linguistic Variability and Intellectual Development*. Translated by G. C. Buck & F. A. Raven. Philadelphia: University of Pennsylvania Press.
- Waterson, N. (1971). Child phonology: a prosodic view. *Journal of Linguistics*, 7, 179-211.
- Wilcox, S. (1992). *The Phonetics of Fingerspelling*. Philadelphia, PA: John Benjamins.