

3 How did language go discrete?

MICHAEL STUDDERT-KENNEDY

'Human language is based on an elementary property that also seems to be biologically isolated: the property of discrete infinity.'

(Chomsky (2000: 3))

3.1 Introduction

'Discrete infinity' refers to the creative property of language by which speakers construct and hearers understand, from a finite set of discrete units, an infinite variety of expressions of thought, imagination, and feeling. This is the property that Chomsky has been endeavouring to describe and explain throughout his career. For Chomsky, the central, 'biologically isolated' property of language and the source of its infinite scope is syntax, the abstract, linguistic, computational system by which discrete, meaningful units (morphemes, words) are combined to form an infinite variety of phrases and sentences. An important mechanism of syntax is recursion, the embedding of a phrase within a phrase that permits, in principle, sentences of infinite length and affords language much of its power to express new thoughts appropriate to new situations. Indeed, in a recent paper, Hauser, Chomsky, and Fitch (2002: 1571) term recursion a 'core property' of language. They defend the hypotheses that (1) 'FLN [the faculty of language in a narrow sense] comprises only the core computational mechanisms of recursion', (2) 'only FLN is uniquely human', and (3) 'most, if not all, of FLB [the faculty of language in a broad sense] is based on mechanisms shared with non-human animals' (2002: 1573).

Hypotheses limiting the 'uniquely human' properties of language to a single syntactic mechanism without regard to phonology or semantics

have no empirical basis, however. The hypotheses derive rather from the purely formal 'architecture', the 'syntactocentric' model of language that generative linguistics has assumed, without evidence or argument, since its inception (Jackendoff 2002: 107-11). Hauser et al. (2002: 1570) illustrate the model in their figure 2 with concentric circles: recursion at the centre, surrounded by 'conceptual-intentional', 'sensory-motor', and other 'broad' processes at the periphery. Such a model does not lend itself to evolutionary interpretation because it suggests no path from sensory-motor primitives to elaborated cognitive syntax by which supporting neural structures might have grown and evolved. In fact, the model is simply one of many possible descriptions of language form. Yet what we need, if we are to understand how language became part of the human genetic endowment, is a model of language function. Form follows function, not function form.

The present chapter attempts to contribute to a model of function by starting from the bottom up instead of the top down. The chapter argues that phonology, though it lacks recursion, has its own mode of discrete infinity (cf. Jackendoff 2002: 111-17), no less 'biologically isolated' than that of syntax, namely, its capacity to form an unbounded set of meaningful words by combining a finite set of meaningless discrete phonetic units. Indeed, it is in words rather than in syntax that the child first evinces, by systematic speech errors, a combinatorial capacity for discrete infinity (see section 3.8, below). Moreover, as Bickerton has remarked, 'Syntax could not have come into existence until there was a sizable vocabulary whose units could be combined into complex structures' (1995: 51). And, we may add, a sizable vocabulary, beyond the 30-40 vocal signals typical of modern primates, could not have come into existence until holistic vocalizations had differentiated into categories of discrete phonetic units that could be organized into words (Lindblom 1992, 1998; Studdert-Kennedy 1987; Wray 2000). Thus, discrete phonetic units are, logically and biologically, necessary precursors of syntax and therefore the ground of Chomsky's discrete infinity.

The combination of discrete phonetic units into a functional hierarchy (gestures, segments, syllables) seems, moreover, to be an example of a biologically unique mode of behavioural organization. Several studies, seeking analogues in animal behaviour, have discovered statistically defined, higher-order groupings of elementary behavioural units in, for example, blowfly grooming (Dawkins 1976), mouse grooming (Fentress

1983), chickadee calls (Hailman and Ficken 1987), and gibbon 'song' (Mitani and Marler 1989). But none has found new levels of structure or function in higher-order units analogous to the structure of a syllable or the meaning of a word. Of course, even if chickadee calls, gibbon 'song', or any other animal communication system were indeed analogues of human language, the prospects for further behavioural studies along these lines would still not be good—at least if we assume for animal systems what Chomsky has long (correctly) maintained for language (e.g. 1965: 21) that, in principle, no objective procedures for discovery of linguistic structure beyond a speaker-hearer's intuition are possible.

Not surprisingly, then, 'duality of patterning' (Hockett 1958), the two-tiered hierarchy of phonology and syntax that affords language its infinite expressive scope, is commonly taken to be a 'fundamental universal characteristic' of language (Hurford 2002a: 319). The origin of this 'universal characteristic' is seldom considered, however (although see Carstairs-McCarthy 1999, and also Tallerman, Chapter 6, for a critique). Certainly, Hauser et al. (2002) have nothing to say on the matter. The purpose of what follows is to suggest how this hierarchy might have begun to evolve. My key assumptions are two: (i) The elements of cognitive structure initially arise *in evolution*, directly or indirectly, from sensorimotor experience: brains evolved to perceive the world and to control action; (ii) functional (articulatory and acoustic) properties of speech give rise to phonological form, not vice versa. I will propose that discrete phonetic units evolved in step with vocal imitation, a capacity unique among primates to humans (Hauser 1996), by differentiation of the vocal apparatus into six discrete, independently controlled organs, and that vocal imitation evolved out of the manual and facial imitation characteristic of an earlier hominid mimetic culture (Donald 1991), perhaps supported by systems of mirror neurons such as those recently discovered in macaques (Rizzolatti and Arbib 1998; Arbib, Chapter 2 above).

3.2 The analogy between language and the genetic system

3.2.1 Meaningless discrete units

That speech can be broken into strings of phonetic units, consonants, and vowels was, of course, known to ancient Greek and Roman grammarians;

the latter, indeed, gave them their names, *littera vocalis* and *littera consonans*, in which letters of the alphabet are explicitly identified with sounds of speech. And, as a matter of fact, until x-rays and magnetic resonance images of speech became available, the transcription of speech by a listener into a string of discrete alphabetic symbols and its recovery by a reader were our only evidence for the standard assumption that we speak by permuting and combining a finite number of discrete elements of sound.

The 'potential infiniteness' of language as 'a system of sound-meaning connections' was recognized, as Hauser et al. (2002: 1571) observe, by Galileo and the seventeenth-century 'philosophical grammarians', among others. Yet, unlike Hauser et al. (2002), these writers saw 'potential infiniteness' as arising, in the first instance, not from the combination of meaningful units or from syntactic recursion (of which they knew nothing), but from the combination of meaningless letters of the alphabet or speech sounds to form words. In passages quoted elsewhere by Chomsky (1966: 84), we find Galileo (1632/1953: 119) attributing the vast scope of written language to 'the various collocations of twenty-four little characters upon a paper', and the 'philosophical grammarians' admiring '...cette invention merveilleuse de composer de vingt-cinq ou trente sons cette infinie variété de mots... n'ayant rien de semblable en eux-mêmes à ce qui se passe dans notre esprit...' ['...this marvellous invention of forming an infinite variety of words from twenty-five or thirty sounds that in no way resemble in themselves what goes on in our mind...']. (Arnould and Lancelot 1660/1997: 23). Here, the authors stress that the basic combinatorial elements they are thinking of are arbitrary and have no intrinsic meaning, a point that Hauser et al. (2002) fail to mention.

Similarly, in the analogy between language and the genetic system with which they open their paper, Hauser et al. (2002: 1569) remark on certain correspondences—hierarchy, generativity, recursion, limitless scope—but ignore the correspondence that makes all the others possible, namely, the intrinsic lack of meaning in both gene and phonetic segment. If meaning, or function, inhered in the elementary units, they could hardly commute across contexts to form larger structures with new meanings or functions. Thus, Jakobson (1970: 438), a linguist, remarked that both 'the genetic code and the verbal system are... based upon the use of discrete components... devoid of inherent meaning', and Jacob (1977: 188), a biologist

writing of language and heredity, stressed that 'for such a system to function implies that the basic units, phonemes or chemical radicals, are by themselves devoid of meaning'. Yet Hauser et al. (2002) propose a model in which the basic mechanism, common to both genetics and language, of combining meaningless units to form a hierarchical structure of meaning (or function) above them, is not even mentioned.

3.2.2 *The particulate principle*

The analogy between language and the genetic system, captured in the now standard textbook metaphor of the 'genetic code' and adopted without question by Hauser et al. (2002: 1569), was a source of puzzlement to Jakobson (1970), one of the first to remark it. He raised the question of 'whether the 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 modelled directly on its structural principles' (1970: 440). The answer, convergence of function rather than homology of form, was implicit in Jacob's observation that the principle of a combinatorial hierarchy '... appears to operate in nature each time there is a question of generating a large diversity of structures using a restricted number of building blocks' (Jacob 1977: 188). But it was Abler (1989) who spelled out the logic of the answer, and so invited a more illuminating and evolutionarily tractable model of language than the syntactocentrism of Hauser et al. (2002).

Abler (1989) was the first to extend to other domains Fisher's (1930) arguments concerning the discrete combinatorial (as opposed to blending) mechanisms of heredity. He recognized that a combinatorial and hierarchical principle is a mathematically necessary condition of all natural systems that 'make infinite use of finite means', including physics, chemistry, genetics, and language. He dubbed it 'the particulate principle of self-diversifying systems'. Briefly, the principle holds that all such systems necessarily display the following properties: (i) Discrete units drawn from a finite set of primitive units or elements (e.g. atoms, genes, phonetic segments) are repeatedly permuted and combined to yield larger units (e.g. molecules, proteins, syllables/words) above them in a hierarchy of levels of increasing complexity; (ii) at each level of the hierarchy, larger units have structures and functions beyond and more diverse than those of

their constituents from below; (iii) units that combine into a larger unit do not disappear or lose their integrity: they can re-emerge or be recovered through mechanisms of physical, chemical, or genetic interaction, or, for language, through the mechanisms of human speech perception and language understanding. (For fuller discussion, see Studdert-Kennedy 1998, 2000.)

For an evolutionary account of language the importance of the particulate principle is twofold. First, it brings language within the natural sciences by generalizing its combinatorial, hierarchical structure across other domains; in other words, *it derives duality of patterning from a broad extralinguistic physical principle rather than accepting it as a language-specific cognitive axiom*. Second, the principle invites a view of language as a hierarchy of increasing scope, and complexity, evolving by stages from basic symbolic reference, or 'naming' (Terrace 1985, 2005), and a combinatorial phonetics, through the simple word combination of a proto-language (Bickerton 1990), to the elaborate combinatorial phrase structures of recursive syntax (cf. Dessalles 2000: part II; Jackendoff 2002: chapter 8). On such a model, each step in the evolution (or ontogeny) of the 'language organ' sets up the structural conditions for the next step.

3.3 The elements of speech and language

3.3.1 *The linguistic status of consonants, vowels, and features*

Whether we acknowledge the particulate principle, as logic demands, or simply accept the dual pattern as an unexplained, language-specific axiom, as does Hurford (2002a), we cannot avoid the question of what the elements of spoken language actually are. Curiously, a common answer, even among writers concerned with the evolution of language (e.g. Hurford 2002a; Kohler 1998; MacNeilage 1998), is that the basic units are consonants, vowels, and their descriptive features.

The answer is curious for at least three reasons. First, consonants and vowels have no existence outside language: they are purely linguistic entities and therefore precisely what an evolutionary account of phonology must explain. No one undertaking an evolutionary account of syntax takes its major word classes, such as nouns and verbs, for granted. They attempt rather to ground the classes in properties and events of the

external world to which early hominids may be presumed to have been sensitive, by invoking such extralinguistic notions as Agent, Patient, Goal (e.g. Bickerton 1990). Similarly, consonants and vowels must have arisen from prelinguistic primate perceptuomotor capacities. How did this come about?

A second reason for rejecting consonants and vowels as basic elements is that they are compound units, analogous to molecules, not atoms. They are commonly said to be bundles or strings of features. But, as the adjectival nomenclature of feature systems reveals (labial, coronal, nasal, etc.), features are descriptive properties, not substantive entities, and necessarily therefore, like the segments they describe, purely linguistic. Features too must derive from prior non-linguistic perceptuomotor capacities.

A third reason for rejecting consonants and vowels is that we have known for over fifty years, since the publication of Joos' *Acoustic phonetics* in 1948, that discrete units corresponding to consonants and vowels are not to be found in the acoustic signal (Liberman, Cooper, Shankweiler, and Studdert-Kennedy 1967). The standard response to this paradox, implicit but not always acknowledged, has been a retreat into cognition. The retreat was sounded by Hockett (1958: 210) in his famous simile likening a string of phonemes passing through the speech apparatus to a row of variously coloured, but unboiled Easter eggs passing through a wringer. On this view, discrete units exist only in the minds of speakers and hearers. Special implementation rules are applied by the speaker and special (or, for some authors, general) perceptual mechanisms are deployed by the listener. But, for an evolutionary account, the retreat into cognition will not do. How, after all, did discrete units get into the mind?

3.3.2 *Insights from reading machines for the blind*

Clues to the true nature of the units of speech came from work on reading machines for the blind, devices designed to transform optic print into an acoustic alphabet—a sort of auditory Braille (Cooper, Gaitenby, and Nye 1984). Despite years of work in several countries, no one succeeded in devising a set of acoustic patterns more efficient than the dots and dashes of Morse code, for which highly skilled professional users may reach reception rates of some 30–40 words/minute, roughly one fifth of a

normal English speaking rate and intolerable for extended listening. Why is speech so much more efficient than any artificial acoustic code? In English, we readily produce and comfortably understand 120–180 words/minute or 10–15 phonetic segments/second. (Readers may want to check these numbers by reading a text out loud at a brisk rate for a minute.) If we break the phonetic segments down into discrete movements of lips, tongue, velum, and larynx, we arrive at a rate of some 15–20 movements/second. By way of comparison, a violinist's tremolo may reach 16 Hz and a hummingbird may beat its wings at over 70 Hz. But these are identical repetitive movements of a single organ. Speech, by contrast, engages half a dozen organs (lips, tongue blade/body/root, velum, larynx) in as many different combinations as there are different phonetic segments in the speech stream, all nicely executed within a tolerance of millimetres and milliseconds. In fact, it is precisely the distribution of action over different articulators that makes the high rates of speech possible.

Such rates 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 entities, 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 (Liberman et al. 1967: 446, italics added).

Here, departing from the purely sequential processing usually assumed for patterns of sound distributed over time, the authors propose both for speech production and, by implication, for speech perception (as many subsequent studies have indeed confirmed) a mode of parallel processing such as enables us to perceive patterns of light distributed over space so rapidly—most notably in reading. The authors see the greater rate of speech compared with arbitrary acoustic alphabets as due to rapid moment-to-moment changes in vocal tract configuration (and so in spectral structure) effected by the orchestrated actions or momentary positions of *all vocal organs simultaneously* (cf. Lindblom 1998: 261). They propose, as a basic phonetic element, a change in position by an articulator, that is, a unit of action. Importantly, unlike the static acoustic alphabets devised for reading machines, the proposed unit is dynamic.

Notice also that, unlike consonants, vowels and features, movements of the vocal apparatus are not intrinsically linguistic. Almost every movement of the articulatory machinery, later refined and adapted to speech,

may be found in crude form in primate cries and calls (glottal action, lip protrusion, mandible lowering and raising) or in primate sucking and feeding (smacking the lips, lowering and raising the mandible for chewing; raising, lowering, humping, and retroflexing the tongue to prepare a bolus of food for swallowing, and so on). Thus, many speech gestures are adaptive variants of prior non-speech units of action (cf. MacNeilage 1998).

3.3.3 *The gesture in articulatory phonology*

Once a movement has been selected for linguistic use, it may properly be termed a gesture, that is, a movement made with communicative intent. The word 'gesture' has been widely used informally in the speech literature for many years to refer to articulatory movements. Only recently has the term been given a precise, technical definition and formally adopted by Browman, Goldstein, and their colleagues, as the central concept in the new and still developing theory of articulatory phonology (Browman and Goldstein 1986, 1991, 1992, 1995).

In that theory the gesture serves as a unit of motoric, phonetic, and phonological function. Motorically, a gesture is the act of forming and releasing a constriction of variable location and degree, at some point in the vocal tract, thus effecting a dynamic sequence of vocal tract configurations. (The reason for including both formation and release of a constriction within a single gesture is the fact, established by many perceptual studies, that acoustic information specifying any given gesture is distributed over the spectral structure both before and after the peak or centre of the gesture itself.) Phonetically and phonologically, a gesture achieves some communicative goal, such as lip closure, velum lowering, tongue raising, glottal abduction, and so on. Thus, the gesture is simultaneously a concrete unit of phonetic action and an abstract unit of phonological contrast.

I will not rehearse here arguments and experimental evidence for the gesture, all readily available in the substantial literature on articulatory phonology and in several recent papers (e.g. Browman and Goldstein 2000; Goldstein, Pouplier, Chen, Saltzman, and Byrd, forthcoming; Studdert-Kennedy and Goldstein 2003). I note only the following points that distinguish gestures from the purely linguistic entities of consonants, vowels, and features. First, as a unit of phonetic action the gesture can

be directly observed by a variety of techniques, including x-ray, magnetic resonance imaging, and palatography. (In fact, it was by copying the movements of discrete articulators from x-ray microbeam data for use in articulatory synthesis of speech by a vocal tract model that the gesture of articulatory phonology was first described and defined.) Second, because the gesture is not intrinsically linguistic, we can trace a continuous path from infant prelinguistic mouthings and vocalizations through babbling and early speech to the mature phonological system (e.g. Nitttrouer 1993; Studdert-Kennedy 2002; Studdert-Kennedy and Goodell 1995). Third, the gesture takes a step toward the desired evolutionary account of the origins of consonants and vowels, and their descriptive features, by viewing them as recurrent, complex, cohesive patterns of gesture and sound. The last point is particularly important because it allows us to see consonants and vowels as dynamic units of phonetic action, spatially and temporally coordinated gestural structures, rather than as static 'beads on a string', as the saying goes.

3.4 Where do gestures come from?

A fair conclusion from the argument up to this point is that the shift from continuously variable primate signalling to particulate human language required an integral anatomical system of discrete, independently movable parts that could be coordinated to effect rapid sequences of expressive global action. The only candidate systems among primates (perhaps, indeed, in the entire animal kingdom) are the hands, the face, and the vocal apparatus. One might readily develop an argument for signing with hands and face similar to what follows for speech with face and vocal apparatus, but I focus on the latter, if only because facial expression and vocalization (together with bodily posture) are the principal means of communication among our closest primate relatives. How then, we must ask, did the human face and vocal apparatus come to be so much more highly differentiated than those of other primates? How did we come to have independent control over the several parts, or organs, of these two systems? And how did we come by the capacity to coordinate the actions of these organs into cohesive facial expressions and vocal tract configurations?

Ultimately, the evolution of such capacities must rest on changes in morphology and neurophysiology. 'The crucial factor, however, in the

acquisition of most evolutionary novelties is a shift in behavior... [C]hanges in behavior generate new selection forces which modify the structures involved' (Mayr 1982: 611-12). My hypothesis is that the 'crucial factor' in differentiation of the human face and vocal apparatus was the evolution of facial and vocal imitation, both behaviours unique among primates to humans (Hauser 1996). How and when did the capacity to imitate first arise?

3.5 Mimetic culture

The gap between the collective habits of ape and human social groups is so wide that we are compelled to posit intervening stages of prelinguistic hominid social organization. By far the most deeply thought-out account of such an intervening culture is Donald's (1991) hypothesis of a mimetic culture in *Homo erectus*, generally believed to have been an immediate predecessor of *Homo sapiens*. *Homo erectus*, a stable species for more than a million years, spread out over the entire Eurasian land mass, leaving evidence of a complex nomadic life well beyond the reach of apes: they made stone tools, practised butchery, and used fire. What held groups of these creatures together? How did individuals communicate with one another? How did they plan group activities? We shall probably never know the answers to these questions, but here I adopt Donald's (1991) well reasoned, though necessarily speculative, account of how the capacity for bodily imitation first arose.

Donald sees the modern human mind as a hybrid of its past embodiments. In the brachiomanual gestures, facial expressions, pantomime and inarticulate vocalizations to which modern humans readily resort when deprived of the use of language, Donald sees a powerful and coherent mode of communication and thought that he terms mimesis. The capacity for mimesis, Donald proposes, first arose in *Homo erectus*. The mode requires conscious and intentional control of expressive behaviours, including facial mimicry and inarticulate vocalization, which we still use and understand in dance, pantomime, and even daily life. We are justified in regarding mimesis as a possible independent mode of communication that evolved before language, because it emerges naturally in deaf and aphasic individuals who cannot speak or use a formal sign language, and

in normal individuals who find themselves among speakers of a language they do not know.

Mimesis is a continuous, analogue, iconic mode of representation, instantaneous in its effect, holistic and idiosyncratic. Language, by contrast, is particulate (or digital), non-iconic, linear and hierarchical, analytic and conventional (grammatical). How did the face and vocal apparatus go from analogue to discrete? How did they differentiate into discrete independently controllable organs? For this we must consider what little we know about the mechanism of imitation.

3.6 How do we imitate?

Imitation is central to human life and culture. Almost everything we do in our daily lives, other people also do. We are creatures of conformity who have learned to live by copying others. Yet we know remarkably little about how we do this. We wave goodbye to a one-year-old infant, in its mother's arms, and we are not surprised when the infant waves back. But how does the child do this? How does it transduce the optic pattern falling on its retinae into the matching pattern of movements in its own hand and arm? How does light get into the muscles?

'I would not be surprised if specific neurons were found that carry out some of the basic tasks of imitation, such as relating observed facial expressions or actions to one's own...' (Blackmore 1999: 80-1). When Blackmore published these words, evidence for such neurons had already been found in the macaque monkey by Rizzolatti and his colleagues at the University of Parma (Rizzolatti, Fadiga, Gallese, and Fogassi 1996).

3.6.1 Mirror neurons

Certain neurons in ventral premotor cortex (area F5) of the macaque brain fire not only when the monkey performs an action, but also when it sees another monkey or another creature with similar gross anatomy, a human experimenter, perform the same action (Rizzolatti et al. 1996; Arbib, Chapter 2). The first actions for which these so-called 'mirror neurons' were reported were manual actions, such as grasping, twisting, tearing, or otherwise manipulating pieces of food. Later studies found F5

neurons (i) for both seeing and producing mouth actions, including communicative mouth actions, such as lip protrusion and lip-smacking (Ferrari, Gallese, Rizzolatti, and Fogassi 2003), and (ii) for both hearing and/or seeing and producing sound-making actions, such as ripping paper, dropping a stick, or breaking a peanut (Kohler, Keyser, Umiltà, Fogassi, Gallese, and Rizzolatti 2002).

According to Rizzolatti and Arbib (1998), area F5 is somatotopically organized: its dorsal part contains a representation for hand movements, its ventral part a representation for mouth and larynx movements. The rostral part of the macaque ventral premotor cortex is believed by many to be the homologue of human Broca's area. Importantly, 'in most F5 neurons, the discharge correlates with an action, rather than with the individual movements that form it' (Rizzolatti and Arbib 1998: 18). In other words, F5 is organized not only *somatotopically*, but also by *function* or *action*. Since macaques are not known to imitate the actions they observe, either in the wild or in the laboratory, Rizzolatti and Arbib (1998: 190) postulated 'a fundamental mechanism for action recognition' in these monkeys.

Evidence consistent with both manual and vocal mirror neuron systems in humans comes from transcranial magnetic stimulation (TMS) studies. TMS, focused on specific cortical sites, can either excite or inhibit the neurons that activate specific muscles of which the response can then be measured by changes in electrical potential associated with muscle contractions, that is, by motor-evoked potentials (MEPs). Fadiga and his colleagues have shown that MEPs recorded from hand muscles are significantly increased over the levels induced by TMS alone, if subjects are simultaneously observing, or even simply imagining, movements normally executed by those muscles (Fadiga, Fogassi, Pavesi, and Rizzolatti 1995). Similarly, listening to words or pseudowords (nonsense) containing medial linguapalatal trills (e.g. Italian *birra* (beer) or *berro* (pseudoword)) significantly increases MEPs, recorded from the tongue muscle that effects the linguapalatal constriction, over the baseline TMS level for that muscle, as compared with listening to words containing medial labiodental fricatives (e.g. *bafo* (moustache) or *biffo* (pseudoword)) or with listening to a pair of non-speech tones (Fadiga, Craighero, Buccino, and Rizzolatti 2002). This result is consistent with a system of mirror neurons underlying both our capacity for vocal imitation and, perhaps, a motor mechanism of speech perception (Studdert-Kennedy 2002).

We should, however, strike two notes of caution. First, the putative speech mirror neurons do not explain how the transform from sound to phonetic action is actually effected; for example, we do not know how they solve the puzzle of the so-called 'inverse transform'. The puzzle arises because, although a given vocal tract configuration gives rise to a unique acoustic spectral pattern, the inverse does not hold: a given spectral pattern may arise from two or more different vocal tract configurations. Of course, the ambiguity may be reduced, or even eliminated, under the dynamic constraints of gestures moving from target to target, so that there would then be no puzzle to solve. Nonetheless, in their perceptual function, speech mirror neurons evidently represent the output from a process of acoustic-to-articulatory transformation that we still do not fully understand.

A second note of caution accompanies the inference from mirror neurons to imitation. For, as remarked above, although macaques (the only creatures in which single mirror neurons have so far been directly observed) recognize, they do not imitate the actions that their mirror neurons represent. On the other hand, recognition that a conspecific shares the same motor repertoire as oneself would seem to be a first and necessary step toward imitation (and, incidentally, toward a 'theory of mind'). We may reasonably hypothesize therefore that mimesis evolved in the hominid line by exploiting mirror neuron systems inherited from primate ancestors. Mirror neurons seem indeed to provide precisely the type of intermodal representation that Meltzoff and Moore (1997) posit in their model of human facial imitation.

3.7 From facial to vocal imitation

3.7.1 Facial imitation in infants

Infants can imitate manual, facial, and vocal actions. Facial imitation is unique among the three modalities because the child cannot feel the face that it sees and cannot see the face that it feels. Facial imitation is therefore necessarily cross- or intermodal.

Much of what we know about infant facial imitation comes from the sustained research programme of Meltzoff and Moore (1997 and many references therein). Among the characteristics of infant facial imitation,

they list the following (1997, table 1): (i) infants imitate a range of specific, isolated, and arbitrary acts, including tongue protrusion, lip protrusion, mouth opening, eye blinking, cheek and brow movements; (ii) infants, presented with a model to imitate, quickly activate the appropriate facial organ (indicating somatotopic representation of facial organs); (iii) infants spontaneously correct their erroneous imitations; (iv) infants imitate absent actions, that is, actions previously, but not currently, observed. These last two characteristics demonstrate that facial imitation is mediated by a representation of the target. We recognize in this list characteristics familiar, *mutatis mutandis*, from infants' early imitations of words.

Indeed, Meltzoff and Moore (1997) propose a model, the active inter-modal matching (AIM) model of facial imitation, that can readily be extended to vocal imitation. Their model is particularly appropriate because it deals not with the 'rational imitation' of function described by Gergely, Bekkering, and Kiraly (2002), for example, but with mimicry of arbitrary facial actions similar to the intrinsically meaningless gestures of vocal mimicry and vocal accommodation (Locke, 1993: chapter 4; Vihman 1996: 115-18). Among the concepts central to AIM are *organ identification*, *body babbling*, and *organ relations*. Organ identification evidently draws on somatotopic representations similar to those by which infants recognize correspondences between their own vocal organs and those of adults. Body babbling, like vocal babbling, is spontaneous activity by which the infant discovers the relations between its movements and the resulting organ configurations. Organ relations (e.g. tongue between lips, tongue protruded beyond lips, eyebrows raised, and so on) are analogous to the vocal tract configurations that the infant learns to recognize in an adult spoken utterance; organ relations are the metric by which infant and adult actions are perceived as commensurate.

3.7.2 The facial-vocal link

Given the importance of facial expression in primate communication (Darwin 1872/1998; Hauser 1996: chapters 4 and 7), and the evidence for mirror neurons responding to and producing communicative mouth actions in macaques, we may reasonably suppose that the capacity for facial imitation, postulated by Donald (1991) for the mimetic culture of *Homo erectus*, would have been supported by a system of facial mirror neurons, such as that of the modern macaque.

We also know that changes in position of the lips, jaw and teeth in rhesus monkeys, as in humans, affect the spectral structure of vocalizations (Hauser, Evans, and Marler 1993). Indeed, the close relation between facial expression and the quality of vocalizations was remarked by Darwin (1872/1998: 96). Thus, we may reasonably hypothesize that, as systematic vocal communication was gradually added to the mimetic repertoire in the transition from *Homo erectus* to *Homo sapiens*, the facial mirror neuron system was gradually coopted and extended to the vocal organs. Thus, the capacity for vocal imitation may have evolved out of the capacity for facial imitation, leading crucially to differentiation of the vocal tract. The end result of this process, as we have seen, seems to be a finely differentiated mirror neuron system for speech in *Homo sapiens*.

3.8 Grounding phonetic categories

3.8.1 The role of imitation in vocal tract differentiation

Imitation has often been dismissed as a factor in language acquisition because the child learning syntax quite evidently does more than imitate: it extracts and applies rules. Nonetheless, imitation is the key to building a lexicon large enough to trigger the onset of word combination and syntax. Imitation of a spoken word requires implicit parsing of the perceived act into its component gestures and their reassembly in correct spatiotemporal sequence (cf. Byrne 2003). We see this quite clearly in the systematic errors of a child attempting its first words (Studdert-Kennedy 2002; Studdert-Kennedy and Goodell 1995). Typically, the child recognizes which articulatory organs to activate, but fails to execute the correct amplitude or the correct relative phasing of gestures. Consider, for example, a twenty-two-month old girl who says ['we:n'di] for ['rezn̩] (*raisin*). She evidently recognized the organs to be activated (lips, tongue-tip, velum), but she omitted the initial tongue-tip retroflexion, while correctly rounding the lips, to give [w] instead of [r]; she correctly closed the tongue tip against the palate, while lowering the velum correctly, but too early, to yield an anticipatory [n], then raised the velum while holding the tongue-tip closure to give a delayed stop, [d], instead of the required fricative, [z]; finally, she released the tongue tip, but delayed opening the glottis, to give an unwanted final vowel, assimilated to the preceding point of tongue-tip

constriction. Such errors indicate that the child has recognized the correspondences between organs of the adult vocal apparatus and its own, but cannot control the amplitude and phasing of its gestures.

In another attempt at [r], word-medial rather than word-initial, the same child offered ['bu'di] for ['bɛri] (*berry*). Here, lip rounding for [r] slides into alignment with tongue raising toward the palate, yielding [u] for [ɛ], and full closure of the tongue tip replaces approximant retroflexion, giving [d] for [r]. Thus, the same combination of gestures for [r] gives rise to different errors in different words. Evidently, the child's target is the word as a whole, and gestures have not yet been fully differentiated from the contexts in which they appear. In due course, repeated use of the same organ in many different contexts leads to independent, context-free control of that organ (Lindblom 1992, 2000). Taking the child's development as an epitome of evolution, we may hypothesize that the six components of the vocal apparatus emerged as independently controlled organs in step with the evolution of vocal imitation. But why was it just these components that emerged?

3.8.2 *Why did the vocal tract differentiate as it did?*

Imitation suggests an account of *how* increasingly fine motor control of already existing primate vocal organs may have evolved, but does not explain *why* the vocal apparatus differentiated into the universal set of six independent organs (lips, tongue tip/body/root, velum, larynx), most or all of which are used in every spoken language. A start has been made toward a perceptuomotor account of vocal tract differentiation by the Distinctive Regions Model (DRM) of speech production (e.g. Carré and Mrayati 1990). The model idealizes the vocal tract as a uniform tube 18 cm long, open at one end, closed at the other. Acoustic theory then defines eight discrete regions of such a tube where deformations, or constrictions, afford greatest acoustic contrast for least articulatory effort. Articulatory effort is minimized by positing that tongue movements are discrete transverse gestures (constrictions), perpendicular to the wall of the vocal-tract, rather than continuous longitudinal movements of the tongue through vocal tract space. Thus, the diphthong [ai], for example, consists of a discrete, narrow constriction by the tongue root or the tongue body in the pharynx, rapidly followed by a discrete narrow constriction of the tongue tip at the front of the palate rather than by a

continuous longitudinal movement of the tongue from back to front. Talkers' systematic use of discrete transverse movements of the tongue has been confirmed through analysis of lateral x-rays by Iskarous (forthcoming). The eight distinct regions of the DRM include all discrete places of articulation known to be used in the world's languages, executed by gestures of the lips and the three organs of the tongue.

3.8.3 *The origins of discrete phonetic units*

Evolutionarily, the initial basis for discrete phonetic units, we may now hypothesize, was differentiation of the discrete organs of the vocal tract. *Discrete organs constitute the universal basis of a particulate phonology in every spoken language.* Not only do they afford a biologically unique speed of serial action, but they also offer a range of contrasting phonetic categories, some or all of which are used in every language. These categories include contrasts in voicing, in nasality and in 'place of articulation' executed by different organs (lips, tongue tip/body/root).

Differences among languages have arisen, due to diverse, unknown historical contingencies, partly from different ways of dividing gestural continua into categories, partly from differences in the degree to which languages have elaborated the gestural structure of their segmental phonetic repertoires (Lindblom and Maddieson 1988).

Different gestural continua afford different phonetic possibilities. Some continua give rise to the abrupt acoustic discontinuities postulated by Stevens's (1989) 'quantal theory': for example, differences in degree of gestural constriction give rise to abrupt switches from laminal to turbulent to interrupted airflow in the vowel-fricative-stop sequence of the English word *eased* ([izd]). Other gestural continua give rise to acoustic continua: for example, the tongue-body front-back continuum for vowels, the tongue-tip continuum for alveolopalatal fricatives (/s-/ʃ/), and the temporal continuum of gestural phasing between different organs in voice onset time (cf. Browman and Goldstein 2000).

Different processes of phonological attunement (or accommodation) among speakers-hearers in different language communities then give rise to different phonological categories along gestural continua, either through mutual vocal mimicry (Browman and Goldstein 2000) or, perhaps, simply through low-level sensory-motor interactions without functional pressures to communicate (Oudeyer, Chapter 4). Importantly, these

studies of attunement indicate how discrete categories may develop along gestural continua that are devoid of natural category boundaries. Categories may emerge as automatic self-organizing consequences of random search through phonetic space and of random interactions among speakers-hearers under certain perceptuomotor constraints (cf. de Boer 2001a, Chapter 5 below; Lindblom 1992, 2000).

3.9 Summary and conclusions

What are the elementary combinatorial units at the base of the two-tiered hierarchy of phonology and syntax that affords language its infinite expressive scope? The traditional answer (consonants, vowels, and their descriptive features) will not do, from an evolutionary point of view, because these are purely linguistic (i.e. cognitive) units and therefore part of what an evolutionary account must explain. How did these discrete units get into the mind?

This chapter proposes a dynamic unit of phonetic action, the gesture, as defined in the developing theory of articulatory phonology, and as observed in x-ray or other images of an active speaker's vocal tract. A gesture is the formation and release of a constriction of variable location and degree produced by one of the six organs of the vocal apparatus (lips, tongue tip/body/root, velum, larynx). A key question for an evolutionary account is: how did these organs differentiate as independently controlled components of the vocal apparatus? The chapter proposes that the organs differentiated in step with the evolution of facial and vocal imitation, both behaviours unique among primates to humans.

The capacity for analogue mimicry of facial expressions perhaps first emerged in the hypothesized mimetic culture of *Homo erectus*, with the support of a system of mirror neurons as observed in modern macaque monkeys. Repeated use of the same facial organs (lips, cheeks, eyebrows, etc.) in many different expressive contexts led to their independent control as discrete components of an expression. The close link between facial expression and the quality of simultaneous vocalizations led to the beginnings of vocal mimicry. Eventually, repeated reuse of the same six vocal organs in many different contexts led to their emergence as independent organs, supported by a system of mirror neurons such as seems to underlie the speech of modern humans.

These six discrete organs, composing an integral vocal apparatus, are the universal basis of discrete phonetic actions, and so of discrete units of meaning, in every spoken language. Differences among languages arise through diverse, unknown historical contingencies from different processes of phonological attunement among speakers-hearers in different language communities, and from differences in the complexity with which languages elaborate the gestural structure of their segments.

ACKNOWLEDGEMENTS

I thank Louis Goldstein and Björn Lindblom for inspiration and discussion, René Carré, Marc Hauser, John Locke, Pierre-Yves Oudeyer, Maggie Tallerman, and two anonymous reviewers for valuable comments. Preparation of the chapter was supported in part by Haskins Laboratories.

FURTHER READING

I have assumed that early evolutionary steps into language entailed differentiation of the primate vocal apparatus and its neural support, both central and peripheral; but surprisingly little is known about the comparative peripheral neuroanatomy of the vocal tract. For some first steps, see Ramsay and Demolin (2002), Sanders (2002, 2004), Zur, Mu, and Sanders (2004).

On the possible role of mirror neurons in the evolution of imitation in humans, see Rizzolatti, Craighero, and Fadiga (2002); for their possible role in the evolution of language see several papers in Stamenov and Gallese (2002).

For discussion of gestural coordination and the emergence of segments within the framework of articulatory phonology, see Fowler (1996); also papers by Browman and Goldstein, cited above.

For moves toward a functional rather than purely formal account of the sound patterns of language, including computational modelling of phonological systems and the emergence of phonetic gesture, see Lindblom (1992, 1998, 2000) and references therein; also, de Boer (2001a), Browman and Goldstein (2000), Carré and Mrayati (1990), and Oudeyer (Chapter 4 below).

References

- Abler, William L. (1989), 'On the particulate principle of self-diversifying systems'. *Journal of Social and Biological Structures*, 12: 1-13.
- Arnauld, Antoine and Lancelot, Claude (1660/1997), *Grammaire générale et raisonnée* (Paris: Editions Allia).
- Bickerton, Derek (1990), *Language and Species*. (Chicago: Chicago University Press).
- Bickerton, Derek (1995), *Language and Human Behavior*. (Seattle, WA: University of Washington Press).
- Blackmore, Susan (1999), *The Meme Machine*. (Oxford: Oxford University Press).
- Browman, Catherine P. and Goldstein, Louis F. (1986), 'Towards an articulatory phonology'. *Phonology Yearbook*, 3: 219-252.
- Browman, Catherine P. and Goldstein, Louis F. (1991), 'Tiers in articulatory phonology, with some implications for casual speech; in John Kingston and Mary E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. (Cambridge: Cambridge University Press), 341-376.
- Browman, Catherine P. and Goldstein, Louis F. (1992), 'Articulatory phonology: an overview'. *Phonetica*, 49: 155-180.
- Browman, Catherine P., and Goldstein, Louis F. (1995), 'Dynamics and articulatory phonology', in Timothy van Gelder and Robert F. Port (eds.), *Mind as Motion* (Cambridge, MA: MIT Press), 175-193.

- Browman, Catherine P. and Goldstein, Louis F. (2000), 'Competing constraints on intergestural coordination and self-organization of phonological structures'. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, 5: 25-34.
- Byrne, Richard W. (2003). 'Imitation as behavior parsing'. *Philosophical Transactions of the Royal Society of London B*, 358:529-536.
- Carré, René and Mrayati, Mohamed (1990), 'Articulatory-acoustic-phonetic relations and modeling, regions and modes'. In Alain Marchal and William J. Hardcastle (eds.) *Speech Production and Speech Modeling*. (Dordrecht: Kluwer Academic Publishers).
- Carstairs-McCarthy, Andrew. (1999), *The Origins of Complex Language*. (Oxford: Oxford University Press).
- Chomsky, Noam (1965), *Aspects of the Theory of Syntax*. (Cambridge, MA: M.I.T. Press).
- Chomsky, Noam (1966), *Cartesian Linguistics* (New York: Harper & Row).
- Chomsky, Noam (2000), *New Horizons in the Study of Language and Mind* (Cambridge: Cambridge University Press).
- Cooper, Franklin S., Gaitenby, Jane and Nye, Patrick W. (1984), 'Evolution of reading machines for the blind: Haskins Laboratories research as a case history'. *Journal of Rehabilitation Research and Development*, 21: 51-87.
- Darwin, Charles (1872/1998), *The Expression of the Emotions in Man and Animals* (3rd edition). (Oxford: Oxford University Press).
- Dawkins, Richard (1976), 'Hierarchical organization: a candidate principle for ethology', in Patrick P.G. Bateson and Robert A. Hinde (eds.), *Growing Points in Ethology*. (Cambridge: Cambridge University Press), 7-54.

- de Boer, Bart (2001), *The Origins of Vowel Systems* (Oxford: Oxford University Press).
- Dessalles, Jean-Louis (2000), *Aux Origines du Langage* (Paris: Hermes Science Publications).
- Donald, Merlin (1991), *Origins of the Modern Mind* (Cambridge, MA: Harvard University).
- Fadiga, Luciano, Craighero, Laila, Buccino, Giovanni and Rizzolatti, Giacomo (2002), 'Speech listening specifically modulates the excitability of tongue muscles: a TMS study', *European Journal of Neuroscience*, 15:399-402.
- Fowler, Carol A. (1996). Speaking, in H. Heuer and S. Keele (eds.) *Handbook of Perception and Action*, Vol. 2. (London: Academic Press), 503-560.
- Fadiga, Luciano, Fogassi, Leonardo, Pavesi, Giovanni and Rizzolatti, Giacomo (1995), 'Motor facilitation during action observation: A magnetic stimulation study', *Journal of Neurophysiology*, 73:2608-2611.
- Fentress, John C. (1983) 'Hierarchical motor control', in Michael Studdert-Kennedy (ed.) *Psychobiology of Language* (Cambridge, MA: M.I.T. Press), 40-61.
- Ferrari, Pier F., Fogassi, Leonardo, Gallese, Vittorio and Rizzolatti, Giacomo (2001), 'Mirror neurons for mouth actions in monkey ventral premotor cortex', *Society for Neurosciences Abstracts*, 27, Program No. 729.4.
- Fisher, Ronald A. (1930), *The Genetical Theory of Natural Selection* (Oxford: Clarendon).
- Galileo, Galilei. (1632/1953), *Dialogue on the Great World Systems, in the Salusbury translation*. (Edited by Giorgio de Santillana). (University of Chicago Press).

- Gergely, György, Bekkering, Harold, and Kiraly, Idiko (2002) 'Developmental psychology: Rational imitation in preverbal infants'. *Nature*, 415:755.
- Goldstein, Louis F., Pouplier, Marianne, Chen, Larissa, Saltzman, Elliot and Byrd, Dani (in press). 'Dynamic action units slip in speech production errors', *Journal of Phonetics*.
- Hailman, Jack P. and Ficken, Morris S. (1987) 'Combinatorial animal communication with computable syntax: Chick-a-dee calling qualifies as "language" by structural linguistics', *Animal Behavior*, 34, 1899-1901.
- Hauser, Marc D. (1996), *The Evolution of Communication*. (Cambridge, MA:Harvard University Press).
- Hauser, Marc D., Chomsky, Noam and Fitch, Tecumseh. (2002), 'The faculty of language: What is it, who has it, and how did it evolve?' *Science*, 298, 1569-1579.
- Hauser, Marc D., Evans, Christopher S. and Marler, Peter (1993), 'The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations'. *Animal Behavior*, 45: 423-433.
- Hockett, Charles F. (1958), *Course in Modern Linguistics*. (New York: MacMillan).
- Hurford, James R. (2002), 'The roles of expression and representation in language evolution' in Alison Wray (ed.), 311-334.
- Hurford, James R., Studdert-Kennedy, Michael and Knight, Chris (eds.) (1998), *Approaches to the Evolution of Language* (Cambridge: Cambridge University Press).
- Hurford, James and Fitch, Tecumseh (eds.) (2002). *Proceedings of the Fourth International Conference on the Evolution of Language*, Harvard University, Cambridge, MA. March 27-30.

- Iskarous, Khalil R. (in press), 'Patterns of tongue movement'. *Journal of Phonetics*.
- Jackendoff, Ray (2002). *Foundations of Language*. (Oxford: Oxford University Press).
- Jacob, François 1977), 'The linguistic model in biology', in Daniel Armstrong and Cornelius H. van Schoonefeld (eds.), *Roman Jakobson: Echoes of his Scholarship* (Lisse: de Ridder), 185-192.
- Jakobson, Roman (1970), 'Linguistics', in *Main Trends of Research in the Social and Human Sciences*, Vol 1. (Paris/The Hague: UNESCO/Mouton), 419-463.
- Joos, Martin (1948), *Acoustic Phonetics*. Language Monograph No. 23. (Baltimore, MD: Linguistic Society of America).
- Knight, Chris, Studdert-Kennedy, Michael, and Hurford, James R. (eds.) (2000), *The Evolutionary Emergence of Language*, (Cambridge: Cambridge University Press).
- Kohler, Evelyne, Keyser, Christian, Umiltá, M. Alessandra, Fogassi, Leonardo, Gallese, Vittorio and Rizzolatti, Giacomo (2002), 'Hearing sounds, understanding actions: Action representation in mirror neurons,' *Science*, 297,846-8.
- Kohler, Klaus (1998), 'The development of sound systems in human language', in J.R. Hurford, et al., 265-278.
- Liberman, Alvin M., Cooper, Franklin S., Shankweiler, Donald P. and Studdert-Kennedy, Michael (1967), 'Perception of the speech code', *Psychological Review*, 74: 431-461.
- Lindblom, Björn (1992), 'Phonological units as adaptive emergents of lexical development', in Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications* (Timonium, MD: York Press), 131-163.

- Lindblom, Björn (1998), 'Systemic constraints and adaptive change in the formation of sound structure'. In James R. Hurford, et al.: 242-264
- Lindblom, Björn (2000), 'Developmental origins of adult phonology: the interplay between phonetic emergents and the evolutionary adaptations of sound patterns', *Phonetica*, 57: 297-314.
- Lindblom, Björn and Maddieson, Ian (1988), 'Phonetic universals in consonant systems', in Larry M. Hyman & Charles N. Li (eds.) *Language, Speech and Mind* (London and New York: Routledge), 62-78.
- Locke, John L. (1993), *The Child's Path to Spoken Language*. (Cambridge, MA: Harvard University Press).
- MacNeilage, Peter F. (1998), 'The Frame/Content theory of evolution of speech production', *Behavioral and Brain Sciences*, 21: 499-511.
- Mayr, Ernst (1982), *The Growth of Biological Thought* (Cambridge, MA: Harvard University Press).
- Meltzoff, Andrew N. and Moore, M.Keith (1997), 'Explaining facial imitation: A theoretical model'. *Early Development and Parenting*, 6: 179-192.
- Mitani, John C. and Marler, Peter (1989) 'A phonological analysis of male gibbon singing behavior'. *Behaviour*, 109, 20-45.
- Nittrouer, Susan (1993), 'The emergence of mature gestural patterns is not uniform: evidence from an acoustic study', *Journal of Speech and Hearing Research*, 36: 959-972.

- Ramsay, Gordon and Demolin, Didier (2002). A comparative study of the sensory innervation of the vocal tract in humans and primates: possible implications for the development of speech motor control. Part I: the palate, in Hurford, James and Fitch, Tecumseh, 98.
- Rizzolatti, Giacomo, Fadiga, Luciano, Gallese, Vittorio and Fogassi, Leonardo (1996), 'Premotor cortex and the recognition of motor actions', *Cognitive Brain Research*, 3: 131-141.
- Rizzolatti, Giacomo and Arbib, Michael A. (1998), 'Language within our grasp', *Trends in Neuroscience*, 21: 188-194.
- Rizzolatti, Giacomo, Craighero, Laila and Fadiga, Luciano (2002). The mirror system in humans, in Maxim Stamenov and Vittorio Gallese, 37-59.
- Sanders, Ira. (2002). Human tongue, pharynx and vocal fold muscles contain slow tonic muscle, a distinct class of muscle that may have evolved for speech, in Hurford, James and Fitch, Tecumseh, 100.
- Sanders, Ira. (2004). The human tongue slows down to speak, in *Proceedings of the Fifth International Conference on the Evolution of Language*, Leipzig, Germany, March 31-April 3, 2004.
- Stamenov, Maxim and Gallese, Vittorio (eds.). (2002). *Mirror Neurons and the Evolution of the Brain and Language*. (Amsterdam and Philadelphia: John Benjamins).
- Stevens, Kenneth N. (1989), 'On the quantal nature of speech', *Journal of Phonetics*, 1/2: 3-45.
- Studdert-Kennedy, Michael (1987), 'The phoneme as a perceptuomotor structure', in Alan Allport, Donald G. MacKay, Wolfgang Prinz and Eckart Scheerer (eds.) *Language Perception and Production* (London: Academic Press), 67-84.

- Studdert-Kennedy, Michael (1998), 'The particulate origins of language generativity: from syllable to gesture', in James R. Hurford, et al.
- Studdert-Kennedy, Michael (2000), 'Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function', in Chris Knight, et al. 161-176.
- Studdert-Kennedy, Michael (2002), 'Mirror neurons, vocal imitation, and the evolution of particulate speech', in Maxim Stamenov and Vittorio Gallese, 207-227.
- Studdert-Kennedy, Michael and Goldstein, Louis F. (2003), 'Launching language: the gestural origin of discrete infinity'. In Morten Christiansen and Simon Kirby (eds.), *Language Evolution* (Oxford: Oxford University Press), 235-254.
- Studdert-Kennedy, Michael and Goodell, Elizabeth W. (1995), 'Gestures, features and segments in early child speech', in Beatrice de Gelder, and José Morais (eds.) *Speech and Reading: A Comparative Approach* (Hove, England: Erlbaum (UK), Taylor & Francis), 65-88.
- Terrace, Herbert S. (1985), 'In the beginning was the "name"', *American Psychologist*, 40: 1011-1028.
- Terrace, Herbert S. (in press) 'Metacognition and the Evolution of Language', in Herbert S. Terrace and Janet Metcalfe (eds.) *Metacognition: the Missing Link*. (Oxford: Oxford University Press).
- Vihman, Marilyn M. (1996). *Phonological Development*. (Cambridge, MA: Blackwell).
- Wray, Alison (2000), 'Holistic utterances in protolanguage: The link from primates to humans'. In Chris Knight, et al.:285-302.
- Wray, Alison (ed.) (2002), *The Transition to Language* (Oxford: Oxford University Press).