Signed and Spoken Language: Biological Constraints on Linguistic Form, eds. U. Bellugi and M. Studdert-Kennedy, pp. 29-40. Dahlem Konferenzen 1980. Weinheim: Verlag Chemie GmbH.

Clues from the Differences Between Signed and Spoken Language

M. Studdert-Kennedy* and H. Lane**
*Queens College, City University of New York, Flushing, NY 11367 and Haskins Laboratories, New Haven, CT 06510
**Department of Psychology, Northeastern University, Boston, MA 02115, USA

Abstract. The formational structures of signed and spoken language are compared in terms of both their phonemes, or primes, and their features. The comparison leads to the suggestion, first, that the two levels of sublexical structure in both languages provide a kind of impedance match between an open-ended set of meaningful symbols and a decidedly limited set of signaling devices; and, second, that while speech draws on a degree of parallel organization to implement a sequential linguistic structure, sign implements a parallel linguistic structure by a partially sequential organization of its gestures. The differences seem to arise because the hands have more degrees of motor freedom than the mouth and/or because the spatial patterns available to sight afford a richer simultaneous structure than the temporal patterns available to hearing.

INTRODUCTION

If we assume that the two modes of communication, speaking and signing, draw on shared cognitive structures, then systematic differences between spoken and signed languages must result from differences in modality, while similarities may reflect either cognitive properties of language or cross-modality invariances in its implementation. It is such invariances - of motor organization, of perception, or of representation in memory - that may constrain the structure of language.

A fundamental discovery of recent years, due to systematic analysis of American Sign Language (ASL) (4,7), is that a dual pattern of syntax and form characterizes signed no less than spoken language. Although a two-leveled structure is often said to be distinctive of human language, its origin and function are seldom discussed. The functional advantages of the one level, syntax, with its powers of unambiguous predication, repeated recursion, and so on, are apparent. As for its origin, it is not inconceivable that syntactic structure evolved by exploiting neural networks already developed for hierarchical control of motor behavior, but this is a matter well beyond the scope of present speculation. The function of the second level, formational structure, is less obvious and its origin may be more amenable to investigation.

Consider a language with a syntax, that is, rules for forming utterances by combining meaningful elements, but with no phonology, no rules for forming meaningful elements from smaller units. Meaningful elements would then be holistically distinct signals, devoid of systematic interrelations. If the lexicon were iconic, its limits would be set by the human capacity to represent - obviously a more severe constraint for acoustic than for visual form - and abstraction would be difficult, if not impossible. If the elements were not iconic but arbitrary, the lexicon would again be limited, because the number of holistically distinct signals that humans can form at a reasonable rate, vocally or manually, and perceive by ear or by eye, is small. (Most vertebrate communication systems dispose of fewer than 40 distinct signals.) Of course, the lexicon could be enlarged by reduplication of elements (the first step toward structure, incidentally), but this would be a cumbersome solution, making, in the end, prohibitive demands on memory. While a modest lexicon does not preclude a productive syntax, and while listeners will submit to a surprising degree of homonymity (3), it is clear that a lexicon adequate to human cognitive demand could not be constructed without recourse to submorphemic structure.

What are the requirements of such structure? Perceptually, they are simple. First, signals must be attuned to psychophysical capacity. Thus, speech sounds are concentrated in the center of the audiogram and visual information during signed communication tends to concentrate around the observer's line of sight - larger signs with more ample movement occur in the periphery of the visual field, while those requiring finer discrimination occur closer to the fovea. Also, boundaries among phoneme or prime categories must be placed at points of adequate discriminability. There is some evidence for the psychophysical determination of at least some such boundaries in speech, although not yet in sign. But the strongest perceptual demand is that the submorphemic units be so compacted that they place minimal demands on short-term storage before lexical access transfers the processing load to syntactic and semantic mechanisms.

From this perceptual demand spring the motor requirements. The signaler must have at his command a rapid and precise peripheral mechanism with enough degrees of freedom for a fair repertoire of distinct gestures. Speed and precision call for a flexible system with a high degree of central neural coordination. Presumably, it is no accident that cerebral localizations of manual control and linguistic function are associated. Manual and vocal systems probably draw on common principles and mechanisms of motor control.

SERIES-PARALLEL DIFFERENCES IN MORPHEME STRUCTURE

From a linguistic perspective, there are obvious differences between the structures of speech and sign. Most salient are the different ways in which they combine their meaningless units (phonemes, primes). Why does speech combine its units in series, sign in parallel? Or, why does ASL not prefer to fingerspell (using arbitrary units unrelated to those of speech), and why does speech not prefer to stack its units in simultaneous bundles?

The most obvious response is to attribute the series-parallel difference to perception and to the differences between sound and light. The distinction is not clear-cut, but sound does entail primarily a temporal, light primarily a spatial distribution of energy. The distinctive gestures of speech and sign seem to be adapted to the medium through which they are conveyed. For example, the spatial distinctions of tongue height among the whispered high-front-vowel /i/, the fricative /s/, and the stop /t/ (as in <u>east</u>) are a matter of a few millimeters, barely perceptible when viewed spatially by X-ray, but highly discriminable when transduced into the temporal array of sound. Similarly, the extensive use of space in sign language reflects the adaptation of the language to the visual medium. Yet, the visual system is clearly comfortable with a sequential display (ASL compounding, infixing, indexing; negative, topic, and aspect marking) and the auditory system readily discriminates among simultaneous properties (tones, nasalization, stress).

Motor as well as perceptual constraints may underlie the series-parallel difference between modalities. Note first that speech is not entirely sequential. Each phone is formed from a roughly "simultaneous bundle" of articulatory features, and each feature is reflected in the signal by at least some more-or-less simultaneous, often spectrally dispersed, acoustic cues. (We use the term "feature" loosely to refer to an isolable property of a gesture, such as tongue root advancement, glottal closure, or velar opening. We are not here concerned with the abstract features of phonology, each of which may be compounded from several articulatory features. We do, however, propose that, in the last analysis, the feature structure of phonology derives from the feature structure of its modality of expression.)

The feature structure of speech is, in large degree, a consequence of the anatomy and physiology of the vocal tract. The active articulators, carrying the major phonetic load (larynx, tongue, jaw, lips, velum), are few, and each has relatively few discriminable states (here again perception impinges). Moreover, none of the articulators can work in isolation; all are engaged (even if only passively), in the production of any single sound. A sizeable repertoire of sound units can therefore only be built by repeated use of the same articulator, and of a particular action of that articulator in more-or-less simultaneous combination with the several actions of other articulators.

To this extent, speech is no less parallel in form than is sign (see below). We might even wonder why features are not the basic meaningless units of speech and phonemes the basic meaningful elements. Single phonemes are indeed used in many languages to fulfill morphemic functions (interestingly, from the point of view of rate, these are often high frequency grammatical morphemes). However, if this were general, spoken languages would be reduced to a maximum of roughly a hundred morphemes. This limit is placed because many combinations of features are excluded: they call either for the same articulators or for incompatible actions by different articulators. We cannot specify exactly how many combinations are possible without knowing the degrees of freedom of the vocal tract - knowledge that awaits a fuller understanding of its motor control. However, we can estimate the upper limit from the maximum number of phonemes found in any single language, and this is roughly a hundred. Thus, limits on the vocal apparatus force speech, first into a featural structure of its units (phonemes), then into concatenation of those units, in order to achieve an appreciable repertoire of semantic elements.

Yet concatenation carries a penalty: neighboring units are formed by the same small set of articulators, and articulators are limited in the rate at which they can switch from one action to another. Here again, the feature structure of speech permits a solution: carry-over of feature values from one phoneme to the next (1). The opening gesture that releases a consonant is itself a property of the following vowel, while the vowel is, in turn, a precondition of the following consonantal constriction. Thus, as one phonetic unit is produced, the unengaged or partially engaged components of a later unit are being activated: in the word <u>bought</u>, for example, lips round for medial vowel, before they open to release the initial labial consonant, and tongue tip rises for final alveolar closure, while its root is still backed and lowered for the preceding vowel. Thus, the fundamental element of spoken language, the consonant-vowel syllable, is formed by the intricate, overlapping gestures associated with both simultaneous and sequential articulatory features.

Pursuing the series-parallel difference, let us apply this line of reasoning to sign language. There would be too few signs, as there would be too few words, if each was holistically different from the next. Similarly, the primes (hand configurations, locations, movements) from which signs are constructed draw on a modest number of articulators with relatively few possible states. There would be too few handshapes if each shape had to be holistically different from every other, too few movements if each movement shared no features with any other, and so on. Thus, we motivate a level of structure below the level of the prime in sign, as in speech.

But now the types of language part ways. The greater degrees of freedom of the signing apparatus and the visual modality allow sign language to transmit its selected combining elements concurrently rather than sequentially. Occasionally, two primes <u>are</u> sequentially adjacent within a sign, like two phonemes in a word. This small set of signs is then subject to severe phonotactic constraints which tend to make the combining elements maximally opposed on major class features. More commonly, sequentially adjacent primes are separated by a morpheme boundary. For both these reasons, we see little sequential coarticulation in ASL. What we find instead is a tendency for simultaneous elements to interact. Movements are reduced, or shifted from arm to wrist, wrist to finger. Handshapes are adjusted to facilitate contact between body parts. For example, the thumb is moved away from its position across the fingers, as in a fist, to permit the knuckles to touch in the two-handed signs MEET and WASH; the index protrudes from the fist, at the second joint, to contact the face at chin or temple in APPLE and ONION, respectively.

However, we should note that these adjustments are not intrinsic to the manual system as the coarticulations of speech are to the vocal apparatus. The unadjusted handshapes or movements are physically possible, without loss in the rate of information transfer, as the mutual adjustments of consonant constriction and vowel opening are not. In other words, the coarticulation effects of sign language are extrinsic variations, analogous to the presence of aspiration in a syllable-initial English /p/ and its absence in an /sp-/ cluster, rather than intrinsic, as in the spectral and temporal variations that accompany the articulation of a particular consonant before or after different vowels. (For the distinction between extrinsic and intrinsic allophonic variation, see (8)).

In short, a comparison of speech and sign leads us to suggest first, that the two levels of sublexical structure in both languages provide a kind of impedance match between an open-ended set of meaningful symbols and a decidedly limited set of signaling devices; and second, that sign transmits the elemental units at both levels in parallel whereas speech transmits phonemes sequentially, features in parallel. This difference seems to arise because the hands have more degrees of motor freedom than the mouth and/or because the spatial patterns available to sight afford a richer simultaneous structure than the temporal patterns available to hearing.

If this account is correct, we may conclude that it is the differences in modality between speech and sign that determine their differences in morpheme structure. Although spoken language may occasionally make lexical distinctions by means of simultaneous variations in, say, spectral structure and fundamental frequency (as in tone languages), for the most part, it is the ordering of elements that specifies the morpheme, so that whatever coarticulatory interleaving may occur, the basic sequence must be preserved in execution. By contrast, again with some few exceptions, ASL does not use the ordering of elements to distinguish morphemes.

SERIES-PARALLEL DIFFERENCES IN EXECUTION

Yet, as we have already suggested, the series-parallel distinction begins to reverse itself when we examine the detailed processes of execution: parallel processes appear in speech, sequential processes in sign. Thus, Fowler ((2), cf. (6)) has argued that coarticulation effects are due not to the spread of features (such as lip-rounding, velar opening, tongue raising) across neighboring segments but to actual simultaneous or coproduction of consonants and vowels. In this view, the neuromuscular synergisms or coordinative structures involved in vowel production are engaged just once at the start of an utterance and then continue to cycle rhythmically with minor adjustments throughout the utterance. On this underlying and relatively slow rhythmic base are superimposed the actions of the distinct and more rapid coordinative structures involved in consonant production. For example, "lip rounding precedes the measured acoustic onset of a rounded vowel, and therefore coarticulates with the preceding consonants... not because the feature [+rounding] has attached itself in the plan to the preceding consonants, but rather because the vowel /u/ is coproduced with them" (p. 61).

This description of articulation as coocurring coordinative motor structures highlights the resemblance between speech and sign production. The stream of signing can be viewed as the result of coordinative motor structures producing cyclical movements of the arms, on which are superimposed fine movements of the wrists and fingers. The cyclical movements are checked by contact with parts of the body or go unchecked. The dance of the arms on the vertical surface of the body resembles the dance of the tongue on the roof of the mouth. Both systems allow interruption of movement to occur when a moving articulator contacts a fixed or a movable articulator. If the distance from the waist to the crown is greater than that from the lip to the pharnyx, the arm is also longer than the tongue, and a long lever is slow to move. If we recall, further, that the proximal stimulus for sign perception is typically some five feet away from the signer, it is not evident that sign has much greater possibilities than speech for simultaneous transmission either motorically or perceptually.

If we suppose then that interfacing speech and sign with their peripheral articulators imposes similar constraints on each, we are led to inquire where in sign language are the temporally organized coproduction effects, such as Fowler's lip rounding example, that we find in speech. If phonemes (feature bundles) are the first level of submorphemic structure and features the second, where are the changes in the feature bundles caused by interleaving one bundle with another, one set of articulatory configurations overlapping another?

Consider the entry in the Stokoe, Casterline, and Croneberg dictionary (7) for the sign translated in English as LATER. The entry indicates that the phonemically distinct tokens of the three sign aspects that combine simultaneously are (for the dominant hand) L-handshape (as in SHOOT), nodding movement (as in YES), and location on the nonspread flat palm of the nondominant hand (as in CERTIFY). Yet when we look more closely at this example, we are tempted to reorganize the data in such a way that what have traditionally been considered phoneme-like primes are viewed instead as morphemes. Not only are the units involved not meaningless, they are also not fully simultaneous. Rather, they are morphemes that have undergone sequencing and rule-governed alterations. First, the base hand is in the common classifier configuration for flat movable objects (BOOK, PAPER, MIRROR). Let us call it //FMO//. Next, the dominant hand has the pointing configuration used for indexing, for two things pointing at each other (OPPOSE, ARGUE), and for designating units of time (WEEK, MONTH). Call it //POINT//. Finally, the pivotal movement may be related to the rotary movement morpheme in, e.g., BICYCLE: //ROTATE//. We have then a sequence, not a parallel set, of three morphemes, not phonemes or primes. The shift in level of analysis brings the sequential structure of the sign into focus. First the //FMO// occurs; then //POINT// which is realized to agree in position and shape (the thumb is extended) with //FMO//. Finally, //ROTATE// is realized with a nodding action to agree with the prior environment. There is substantial temporal overlap: //POINT// and //FMO// are partly concurrent in execution and move toward agreement in location, orientation, and type of contact. The realization of these morphemes leads to an interleaved sequence of meaningless smaller units including the handshapes, /B,L/ and the movement /ŋ/. Thus, just as analysis of spoken sequence leads to a view of speech as in some degree parallel in its execution, so an analysis of signed simultaneity leads to a view of signs as in some degree serial.

We should emphasize that although the sequential structure of a sign has come into descriptive focus from a reanalysis of its posited prime set as a morpheme sequence, the description does not depend on that reanalysis. (Nor is this the place to propose the general recasting of ASL linguistic structure that this analysis implies.) Rather, the sequencing is entailed by the motoric dimensions themselves. In rapid signing, movement toward location must begin before complete formation of handshape, if location is not to be anomalous; and, if movement is not to be anomalous, handshape and location must be more or less fully established before sign-internal movement begins. In other words, a sequential structure seems intrinsic to sign formation, as a parallel structure is intrinsic to the spoken syllable.

CONCLUSION

We are led to the paradoxical conclusion that sign language draws on a degree of sequential organization to implement a parallel linguistic structure, while speech does precisely the reverse.

But the paradox weakens if we see the two motoric modes as answers to the same communicative demand. The demand is for fluent discourse at a cognitively comfortable rate. The two languages then draw on the same linguistic competence and a common system of central motor control to meet this demand. Their solutions differ in emphasis because they deploy peripheral articulatory structures that differ in their degrees of freedom and that address different perceptual systems.

Acknowledgment. We gratefully acknowledge J. Kegl's important contributions to this paper. Preparation of the paper was supported, in part, by NICHD Grant No. HD 01994 to Haskins Laboratories, and by a grant from the Division of Behavioral and Neural Sciences, National Science Foundation, to Northeastern University.

REFERENCES

(1) Cooper, F. 1972. How is language conveyed by speech? In Language by Ear and by Eye, eds. J.F. Kavanagh and I.G. Mattingly. Cambridge, MA: M.I.T. Press.

(2) Fowler, C.A. 1980. Coarticulation and theories of extrinsic timing. J. Phon. 8: 113-133.

(3) Klima, E. 1975. Sound and its absence in the linguistic symbol. In The Role of Speech in Language, eds. J. Kavanagh and J. Cutting. Cambridge, MA: M.I.T. Press.

(4) Klima, E., and Bellugi, U. 1979. The Signs of Language. Cambridge, MA: Harvard University Press.

(5) Lane, H.; Boyes-Braem, P.; and Bellugi, U. 1976. Preliminaries to a distinctive feature analysis of handshape in American Sign Language. Cog. Psychol. 8: 263-389.

(6) Öhman, S. 1966. Coarticulation in VCV utterance: spectrographic measures. J. Acous. Soc. Am. 39: 151-168.

(7) Stokoe, W.; Casterline, D.; and Croneberg, C. 1965. A Dictionary of American Sign Language. Washington, DC: Gallaudet College Press.

(8) Wang, W.S., and Fillmore, C. 1961. Intrinsic cues and consonant perception. J. Speech Hearing Res. 4: 130-136.