SPEECH SOUND CATEGORIZATION BY CHILDREN

Paula Menyuk, Applied Psycholinguistics, Boston University, Boston, Mass., USA

Clearly acquisition of the structural properties of language takes place via a process of segmenting and then categorizing the segments of the language heard into units which can be used to comprehend and generate unique utterances. Equally clearly, the first aspect of language that is used by the hearing adult and by the infant to segment and categorize utterances is the acoustic speech signal. Unlike the adult, however, who has already determined what the "appropriate units" are and can by-pass much of the surface structure of the utterance, the infant must rely heavily on the signal to come to conclusions about appropriate segmentations. She must also rely heavily on contextual cues to relate the segments of the signal to objects and events in the environment. Despite the fact that common sense tells us that the above must be the case, we are, at the present time, still unclear about what these segments are and what the bases for categorization of segments are either initially or over time as the child matures. Indeed, controversies still exist in the literature on this issue for the adult as well as for the child. In this paper varying hypotheses concerning the nature of and bases for speech sound categorization by the child and its role in language acquisition will be examined in light of theories on adult language processing and the data on the speech processing behavior of the infant and young child.

Theoretical Descriptions of Processing

The segmentation of continuous speech has been described by linguists as being hierarchical and nested. That is, a message can be characterized as a type of speech act. The message can contain a sentence or sentences and these contain phrases which are made up of morphemes. Morphemes are composed of syllables which are made up of speech sound segments each of which represents a bundle of features. If this were a psychologically real description of language processing as well as of elements of language then the listener would determine categories of segments in a sequence with each lower step of the sequence dependent on the immediately higher step since higher steps indicate units of analysis. Speech sound segment categorization would take place at the end of the sequence and require resolution of the bundle of features. Varying descriptions based on this hierarchical model have been labelled "analysis by synthesis" (Cooper, 1972). It might be logically argued from this model that since speech sound categorization or identification is comparatively late in the sequence of on-line processing then it must also be late in the sequence of acquisition of the structural properties of the language.

The above model of language processing has been deemed inadequate in accounting for either real-time processing of speech by the adult or for the observed sequence of acquisition of structural properties by the child. Since earliest utterances are sequences of speech sounds marked prosodically and the infant does not appear to understand anything more about utterances than their affective intent, it cannot account for behavior during the early babbling period. The model does not account for subsequent language behavior since even then the child does not evidence any knowledge of any of the postulated higher categories (i.e. sentence, phrase and morpheme). An alternative description of both processing and the sequence of acquisition is a bottom-up model or "synthesis by analysis". With this model speech sounds are differentiated and categorized, then grouped into higher level categories in a sequential manner during speech processing. In acquisition, speech sounds are differentiated and categorized by a process of imitation and sound approximation which is rewarded. These sounds are then composed into words by the same process and by associating phonological sequences with objects and events. Larger units of an utterance, phrases and sentences, are composed by putting together smaller units via a chaining process (Staats, 1971). This description suggests that the earliest analysis in processing and the earliest structural acquisition are segmental speech categories although the nature of these categories is not defined in the model.

Not only is there a substantial amount of evidence to indicate that this model does not adequately describe adult language processing (Fodor et al., 1974), but, also, it is difficult to see how the analysis of speech sounds one by one in a sequence can lead to decisions about where crucial boundaries lie and, thus, to a determination of meaning. For these same reasons a synthesis by analysis model seems inadequate in accounting for language acquisition. Although there is evidence that in early care-giver-child

communicative interaction, segments and boundaries are made much more salient than they are in adult-adult communication (Newport, 1976), there is no evidence that the child, in the process of acquisition, adds sequentially to segments by chaining bits together or that the child merely imitates input structures. On the contrary, the child appears to be only able to attend to and generate certain aspects of utterances at certain periods of development regardless of input structure and these aspects are not sequential bits of adult utterances.

Still another description suggests that perception and generation of connected speech is a parallel process; i.e. one involving all components of the language simultaneously. In the process chunks of the message, probably phrases, are subjected to analysis and rough estimates are made of the phonological composition of the morphemes in the phrase to corroborate hypotheses about the meaning of the phrase and then other phrases if more than one is contained in the utterance. An exact representation of the phrase can be kept in mind until analysis is completed so that needed corrections on this estimate can be made (Garrod and Trabasso, 1973). The child, during the process of acquisition, would analyze the data in the same fashion. The distinctions between the child and the adult are in the amount of information chunked for analysis, the much heavier reliance on the part of the child on contextual cues for analysis and the process of chunking itself since segmentation strategies would change as more structural knowledge of the language is acquired (Menyuk, 1977, Chap. 5). For example, an early chunking strategy might be to ignore everything in the signal except those sequences that signal main relations of actor and action or action and object. Components of the relation would be grossly analyzed for lexical look-up. However, analysis of the phonological segments per se would not be needed for comprehension. Since the parallel processing requires analysis of segments only when correction of rough estimates is required it might, again, be logically argued that speech sound segment categorizations would be later acquisitions than morpheme categorizations.

The above are theoretical descriptions of adult language processing and theoretical descriptions of the process of language acquisition. In conjunction with these are descriptions which are concerned with phonological acquisition only. This acquisition has

been described as a process of first discriminating between the speech sounds of the language, then categorizing these distinctions in terms of articulatory gestures. These discriminations are based on distinctive feature differences between speech segments. Early distinctions are determined by feature detectors that are preprogrammed in the auditory system of the human infant (Eimas, 1974). These might be termed primary features. Finer distinctions are then made both in perception and production and are probably affected by particular language experience. However, given the universality of the speech processing abilities of normal infants, both perceptually and productively, there is, to some extent, universality in the sequence in which distinctions are made. This universality is confounded by the particular data the child is confronted with; i.e. the language of the child's community and even family. Thus, the universal order is modified by the perceptual and productive problems a particular language poses for the child and by the interactive styles, lexical selections, etc. of a particular family. Individual differences become more marked when standard lexical items in a particular language begin to be used.

Data on Early Speech Processing

On the face of it there appears to be a logical gap between theories of language processing and of language acquisition and theories of the development of the phonological system. The latter suggest very fine analysis of the signal on the segmental level in terms of distinctive features, whereas the former suggest rather gross analyses dependent on higher level categories. There are also large differences between the findings of studies carried out at different periods of early speech processing behavior. One of the primary reasons for these gaps between theories of language and speech acquisition and between the findings of studies of speech processing and the conclusions drawn from them may be not maintaining a clear distinction between what the infant and child can do and what they ordinarily do; i.e. a capacity versus performance distinction. The data collected thus far on early speech processing indicate that the very young infant (1 to 4 months) as well as the very young child (under two years) can discriminate between speech sound segments that vary in terms of a single distinctive feature. There also appears to be a hierarchy in the features that can be distinguished both perceptually and productively. Thus, during

the cooing and babbling periods some features appear to be more perceptually salient than others and this also appears to be the case when the task is distinction of minimal pair nonsense syllables. Similarly, segments containing certain features are realized before segments containing other features in babbled utterances and then in morpheme production. However, there is not an exact correlation between the order of perceptual and productive distinctions made, and individual differences in the exact sequence of features and segments distinguished can be observed.

What seems to be suggested by these data is that distinctions can be made on the basis of distinctive features by the infant and young child if the question is put to them in a way in which these distinctions are made clear; i.e. in a small enough context that is non-distracting such as nonsense-syllables. Also, the response required in the task must be part of the children's behavioral repertoire. For example, they must have sufficient memory to recall the stimuli presented. Finally, there are some features that can be distinguished before others. However, discrimination between features does not imply that categorization of segments has taken place in terms of bundles of features nor does the capacity to discriminate between features imply that this is what children do when they listen to speech and attempt to match articulatory outputs to stored representations. Indeed, all the data indicate that during the babbling and early lexical acquisition periods distinctive feature differences are not actively employed in determining meaning of utterances or in generating utterances.

During the babbling period perceptual processing of continuous speech seems to be primarily based on the supra-segmental aspects of the speech signal and contextual cues. Some time toward the end of this period recognition of a small set of lexical items is observed and still later production of word approximations begins. The lexicon of the child at this time is quite small. It is entirely reasonable to suppose that both lexical recognition and generation are based on syllabic representations of morphemes. In other words, speech processing is taking place on the basis of the morpheme and this may be the minimal unit for categorization of speech information. The meaning of a phonological sequence, its gestalt phonological representation as a syllable or reduplicated syllables, supra-segmental features of intonation and contextual cues appear to be all that is needed or used to comprehend or generate utterances during this time (Menyuk and Menn, in press).

Again, this is what children appear to do in on-line processing of speech during these early periods of development, although, at this time and long before, they are capable of discriminating between speech sounds on the basis of feature distinctions. As the lexicon grows and as structural knowledge increases constraints on memory probably make segmental differentiation and categorization necessary. When this occurs an available competence is actively employed. However, segmental differentiation and categorization may be needed only rarely to comprehend continuous speech. Thus, although the ability may be increasingly used at later periods of development it still may be used infrequently. Research shows that even 3 and 4 year-old children first use morpheme information to differentiate between phonological sequences and only use segmental information with some exertion when morpheme information is unavailable; i.e., with nonsense syllables or unknown words. At present, little is known about when reference to segmental information is used without marked exertion. Such ability is, of course, required in learning to read alphabetic text. One would assume that this ability develops gradually and that there would be individual differences or group variations due to language experience in the ages at which this ability manifests itself (Savin, 1972).

Conclusions

The theoretical description of the processing of language which appears to most adequately describe the sequence of acquisition of the structural properties of the language and to best fit the data on infants and young children's speech processing is one of parallel analysis of chunks of continuous speech. Initially the chunks the child can process are short in duration, linear in arrangement and involve primarily surface structure information. Reference is made to gestalt representations of surface acoustic information to derive meanings. Thus, the analyses are quite gross. As the child matures the chunks that can be processed simultaneously at all levels (semantic, syntactic and phonological) increase in duration and, as structural knowledge grows, recursiveness within chunks can be processed and the analysis becomes more detailed or differentiated. The speech signal must be held in mind and represented to allow analysis using whatever structural knowledge is available. It has been suggested that this representation

or categorization of speech is initially acoustic images of morphemes and/or syllables and only later in terms of segments and features of segments. This appears to be the case even though the infant is capable of discriminating between minimally different acoustic features. In summary, the model that appears to be most descriptively adequate is not a "top-down" or "bottom-up" model but, rather an "outside-in" model (Menyuk, 1977).

References

Cooper, F. (1972): "How language is conveyed by speech", in Language by ear and by eye, J. Kavanagh and I. Mattingly (eds.), 25-46, Cambridge, Massachusetts: MIT Press.

Eimas, P. (1974): "Linguistic processing of speech by young infants", in Language perspectives: acquisition, retardation and intervention, R. Schiefelbusch and L. Lloyd (eds.), 55-74, Baltimore: University Park Press.

Fodor, J., T. Bever and M. Garrett (1974): Psychology of Language, New York: McGraw-Hill.

Garrod, S. and T. Trabasso (1973): "A dual memory information processing interpretation of sentence comprehension", JVLVB 2, 155-167.

Menyuk, P. (1977): Language and maturation, Cambridge, Massachusetts: MIT Press.

Menyuk, P. and L. Menn (In press): "Early strategies for the perception and production of words and sounds", in P. Fletcher and M. Garman (eds.), Cambridge, England: Cambridge University Press.

Newport, E. (1976): "Motherese: the speech of mothers to young children", in Cognitive theory: Vol. II, N. Castellan, D. Pisoni and G. Potts (eds.), Hillside, New Jersey: Lawrence Erlbaum Assoc.

Savin, H. (1972): "What the child knows about speech when he starts to learn to read", in Language by ear and by eye, J. Kavanagh and I. Mattingly (eds.), 319-326, Cambridge, Massachusetts: MIT Press.

Staats, A. (1971): "Linguistic-mentalistic theory versus an explanatory S-R learning theory of language development", in Ontogenesis of grammar, D. Slobin (ed.), 103-152, New York: Academic Press.