Word Segmentation: The Role of Distributional Cues

JENNY R. SAFFRAN, ELISSA L. NEWPORT, AND RICHARD N. ASLIN

Department of Brain and Cognitive Sciences, University of Rochester

One of the infant's first tasks in language acquisition is to discover the words embedded in a mostly continuous speech stream. This learning problem might be solved by using distributional cues to word boundaries—for example, by computing the transitional probabilities between sounds in the language input and using the relative strengths of these probabilities to hypothesize word boundaries. The learner might be further aided by language-specific prosodic cues correlated with word boundaries. As a first step in testing these hypotheses, we briefly exposed adults to an artificial language in which the only cues available for word segmentation were the transitional probabilities between syllables. Subjects were able to learn the words of this language. Furthermore, the addition of certain prosodic cues served to enhance performance. These results suggest that distributional cues may play an important role in the initial word segmentation of language learners.   © 1996 Academic Press, Inc.

Much of the discussion in the field of language acquisition focuses on the nature–nurture dichotomy, and particularly on the ''nature'' side of the equation: Are there innate constraints on the ability to learn language? What is the character of these constraints? While the evidence for some type of innate constraints is extensive, researchers of all theoretical perspectives acknowledge that a substantial part of the acquisition problem must involve learning from the input language. For example, even on a strong nativist/universals view, one part of this learning must entail organizing the input sound sequences into the abstract linguistic categories (e.g., noun and verb classes, as well as higher level phrases) in which linguistic universals are stated; another

part of this learning must involve determining the many arbitrary features which differ quite widely across languages (e.g., what are the sound sequences in this language which form morphemes and words?).

Until recently, surprisingly little attention has been devoted to documenting the course of such learning in language acquisition, or to formulating theoretical proposals about the learning mechanisms which are employed. In the last few years, however, several investigators have begun to consider the learning problem from the point of view of the rich and rather complex acoustic and statistical properties of language which might be stored and computed (cf. Brent, in press, and Morgan & Demuth, 1996, for discussions). In addition, investigators have begun to articulate the surprisingly early abilities of infants and children to appreciate the structure of their particular input language.

Recent research from several labs indicates that children solve many complex learning problems during their first year of life, as they attune their speech perception capacities to the sound structure of their native language. For example, 9-month-old infants prefer to listen to certain prosodic and phonotactic patterns of their native language over those of other languages, whereas 6-month-olds do not (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Fried-

erici, Wessels, Svenkerud, & Jusczyk, 1993). These distinctions are quite fine-grained: 9-month-olds prefer phonetic patterns which occur frequently in their native language over patterns which are also phonotactically legal but which are relatively infrequent (Jusczyk, Luce, & Charles-Luce, 1994). These preferences presumably reflect the infant's ability to monitor the relative frequencies of the prosodic and phonotactic patterns exemplified in their native language. Most importantly for the present discussion, these results reflect learning from the language environment, as the pattern of results differs for infants exposed to different languages.

A somewhat more difficult learning problem is seen in the developmental decline of infant sensitivity to nonnative language phonetic contrasts. Unlike younger infants, year-old infants only discriminate the phonetic contrasts exemplified in the language they are learning (e.g., Best, 1993; Werker & Tees, 1984). This loss of sensitivity occurs even for nonnative phonetic categories which are present in the native language input but which are not contrastive. Such learning requires the ability to determine which exemplars are members of the same phonetic category and which are not, via their distribution in the input. Similarly, Kuhl and her colleagues have demonstrated that the vowel categories of six-month-old infants are organized around prototypes which differ as a function of the vowel structure of the native language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Once again, the structuring of the infant's speech perception presumably reflects the distribution of exemplars in the speech input.

Another complex type of learning which must in large part be input-driven is word segmentation. Unlike the spacing provided in written text, spoken words are rarely bounded by pauses (Cole & Jakimik, 1980), so the infant must somehow discover the boundaries between words in the language input. Because the structure of words is so variable across languages, it is difficult to imagine how innate knowledge could solve this problem. Rather, word segmentation must be accomplished at least in part by an interaction between the child's perceptual and learning mechanisms and the language input. The present paper investigates one possible solution to the word-segmentation problem: distributional cues to word boundaries.

Several different classes of solutions to the word-segmentation problem have been proposed in the literature. One hypothesis is that the child avoids the word-segmentation problem altogether by learning those words which are presented in isolation. This idea is akin to the Bloomfieldian definition of words as ''minimum free forms,'' the smallest units able to constitute individual utterances (Bloomfield, 1933). This solution is potentially problematic as there are many words which are ungrammatical in isolation, notably articles and other function words. Consistent with the Bloomfieldian type of account, however, the closed class generally appears later than the open class in children's language productions and might be discovered only after some word boundaries are initially determined via presentation of content words in isolation (although see Gerken, Landau, & Remez, 1990, for evidence that 2-year-olds who do not produce unstressed function words do in fact perceive them).

However, the weight of empirical evidence runs counter to the words-in-isolation hypothesis. This solution to the word-segmentation problem places the burden of segmentation on the speakers generating the child's language input. One would therefore expect speakers to use new words in isolation whenever possible in their infant-directed speech. In fact, some mothers do not use words in isolation even when explicitly directed to teach their 1-year-old child a new word (Woodward & Aslin, 1990). However, mothers do tend to place new information utterance-finally (even occasionally resulting in ungrammatical structures), pointing to a role for post-utterance pauses in word-segmentation (see Golinkoff & Alioto, 1995, for evidence that adult learning of new words in a foreign language is also facilitated by utterance-final position).

Another possibility is that prosodic cues, such as the post-utterance pauses just mentioned, may play a role in the discovery of

word boundaries. For example, Gleitman and Wanner (1982) proposed that infants might initially consider stressed syllables to be words. More generally, Echols (1993) has suggested that the word-segmentation problem might first be approached by an extraction mechanism: perceptually salient units, such as stressed and certain acoustically distinctive final syllables, might be extracted and identified as initial words. Indeed, mothers of preproductive infants spontaneously raise the pitch of their speech to highlight topical words (Fernald & Mazzie, 1991), and the early (and often incomplete) productions of English-speaking infants consist predominantly of those syllables which are stressed and/or word-final in adult speech (e.g., Echols & Newport, 1992). This perceptual strategy would presumably be quite helpful in initially breaking up the speech stream, although the problem of correctly determining the structure and boundaries of many multisyllabic words remains.

Prosodic cues might also be useful in word segmentation as markers of language-specific word patterns. Some languages contain prosodic cues which are correlated with word boundaries, including word-initial strong syllables in English and word-final accent in French. Cutler, Mehler, and their colleagues have hypothesized that infants are sensitive to these patterns and are able to use them to segment the speech stream (e.g., Christophe, Dupoux, Bertoncini, & Mehler, 1994; Cutler, 1994). However, it is unknown whether all languages provide reliable prosodic cues to word boundaries; with respect to stress, for example, only 211 of the 444 languages surveyed by Hyman (1977) had fixed stress either word-initially or word-finally. Moreover, in order to have learned language-specific prosodic patterns of this type, the learner must also have some knowledge of word boundaries independent of prosodic cues. Without such information, there would be no way to determine whether, for example, stress consistently corresponds to the beginnings of words as opposed to the middles or ends of words. The role of prosody will be considered further in the discussion of Experiment 2.

Another possibility is that children learn words by detecting which sounds co-occur with entities in the environment (e.g., Osgood & Sebeok, 1965). That is, the child may notice that *cat* is often uttered in the presence of the family pet. However, the relationship between the semantics of the world and the child's linguistic input is notoriously complicated (Gleitman, 1990; Gleitman & Gillette, 1995; Quine, 1960). Moreover, 8-month-old infants are able to segment word-like units from continuous speech presented in the laboratory without the benefit of any nonlinguistic information (Jusczyk & Aslin, 1995), and 11-month-olds are able to recognize familiar words without any visible referent (Hallé & de Boysson-Bardies, 1994).

While all of the aforementioned types of information are likely to be quite helpful, they may not always be present in the input or accessible to the learner. There is, however, another type of information which is always present in the input, but which researchers have generally considered to be too complex for infant learners to use: distributional cues (Brent & Cartwright, in press; Goodsitt, Morgan, & Kuhl, 1993; Hayes & Clark, 1970). Words may be defined in distributional terms, along the lines of other syntactic constituents (e.g., Lyons, 1968). A word is a sequence of phones which exhibits positional mobility: if a sequence is picked up and moved to another position in the sentence, it will remain the same word. Words are generally uninterruptable: extraneous information, such as pauses or other words, is generally placed between words rather than in the middle of words.[1] Words also exhibit internal stability: each lexical item consists of a sequence of phones in a fixed order—thus, when the sequence is altered, a new word (or nonword) results.

While words may be defined as fixed sequences of phones, the learner does not have direct access to this information. Rather, what the learner experiences in the input is complex

---

[1] Important exceptions to this generalization include the infixation processes found, for example, in Semitic languages.

statistical information over a corpus of utterances resulting from the concatenation of subword units. This information will take the form of relatively strong correlations between sounds found within words, contrasted with weaker correlations across word boundaries. On this view, one might discover words in linguistic input in much the same way that one discovers objects in the visual environment via motion: the spatial–temporal correlations between the different parts of the moving object will be stronger than those between the moving object and the surrounding visual environment (e.g., Sejnowski & Nowlan, 1994).

The idea that words may be identified as clusters of co-occurring sounds dates back at least to Harris (1955), who proposed that linguists working on unfamiliar languages might discover the individual word units by counting how many phones can succeed a given string. If a large set of possible phones can follow a string, then it is likely that the end of the string is the end of a word. If, however, only a relatively small set of phones is possible, then it is likely that the end of the string is not the end of the word. The size of the successor-counts corresponds to how constraining the phonemic context is; the more constrained, the higher the correlations between phonemes.

On analogy with this discovery procedure, Hayes and Clark (1970) proposed that initial word segmentation by language learners might be achieved using a clustering mechanism measuring crude correlations between segments. Such a mechanism could identify highly correlated clusters of sounds as words, while weaker correlations would indicate word boundaries. In a pioneering study, Hayes and Clark (1970) used a miniature artificial speech stream to test this hypothesis. After synthesizing a set of nonlinguistic sounds (glides and warbles) to represent ''phonemes'', they concatenated these sounds into four words and the words into a speech stream. The speech stream contained no pauses or other cues to word boundaries save the distribution of the phonemes.

Subjects listened to the artificial speech stream for 45 minutes and were then tested on their knowledge of the words. The results provided some indication that subjects were indeed able to recognize strings of words as opposed to strings of nonwords. Moreover, a second experiment demonstrated that decreasing the number of phonemes in the language, and thus reducing the uniqueness of the phoneme combinations found in the words, impeded performance.

While these results are both important and suggestive, there are several issues requiring further exploration. First, Hayes and Clark (1970) found somewhat weak, though significant, results, raising the possibility that learners' abilities to acquire purely distributional information might be limited. Subjects in their experiment scored only 62% correct on a two-alternative forced-choice test. Moreover, the structure of their test was such that even these results might be in part the outcome of test strategies rather than distributional learning. Each test trial consisted of two strings: the four words in random order, separated by pauses (e.g., aaa bbb ccc ddd), and the same four words with each pause shifted two phonemes to the left (dda aab bbc ccd). Thus, on each of the 40 trials, one choice was always the same four words, though in different orders. The second choice varied as a function of the order of the words in the first choice, with a total of 24 possible sets of nonwords. It is therefore possible that subjects succeeded on this test not by virtue of learning during the experimental exposure, but rather by choosing the set of words that was the most familiar by virtue of its presence on every trial of the test. Despite this potential limitation, their experiment raised important and interesting questions which, surprisingly, have received very little subsequent study and which therefore merit further investigation.

On a more theoretical note, Hayes and Clark (1970) leave open the specific nature of the proposed clustering mechanism. If learning is able to proceed based on tabulating the statistics of the input, then it would be worthwhile to further elucidate the form in which those statistics are computed by learners. Hayes and Clark (1970) are not explicit on this issue, instead offering a subjective view

of the learning process based on their experience as pilot subjects: ''after perhaps a minute of listening, an event...stands out from the stream...the listener may notice that it is typically preceded or followed by another event... Recognition of a word, then, seems to proceed from perceptually distinctive foci outwards in both directions toward the word boundaries'' (pp. 227–228). While this description is highly suggestive of the possible structure of the clustering mechanism, further research is required to probe the nature of this learning device.

We therefore endeavored to formalize the idea of the clustering mechanism using a statistical measure: transitional probability. A simple example will illustrate the intuitions underlying the relationship between transitional probability and words in a natural language. Note that while we present this example in terms of sequences of *syllables,* the logic (and therefore the possible psychological mechanisms used by actual language learners) will work as well in terms of sequences of phonemes, features, or other types of sublexical units. Consider the word *baby. Bay* is a fairly frequent syllable in English, which will be followed by *bi* some percent of the time. But *bay* may also be followed by a number of other syllables which it precedes in other words, such as *ba.con, ba.sic, ba.ker,* and *ba.sil.* Each of these syllables will follow *bay* with some probability, depending upon how frequent these words are in the language. These syllable pairs are word-internal, as they do not span a word boundary.

Bay can also occur at the ends of words, as in *obey* and *bay.* In such cases, the syllable *bay* is followed by the first syllable of the next word; in fact, it may be followed by most, if not all, syllables which can begin a word in English. Thus, there exist word-external pairs including *bay#too, bay#sit, bay#me,* and *bay#cal,* each of which spans a word boundary. Over a corpus of English, the word-internal pairs will tend to occur more frequently than the word-external pairs, which are relatively unconstrained. Intuitively, an occurrence of *bay* is more predictive of the following syllable

when both syllables belong to the same word than when the pair spans a word boundary.

This notion may be formalized using transitional probability. The transitional probability of Y given X is computed as follows:

$$\frac{\text{frequency of pair } XY}{\text{frequency of } X}. \qquad [1]$$

A high transitional probability indicates that the presence of X strongly predicts that Y will occur. Lower transitional probabilities signal a weaker contingency between X and Y.

Returning to the example above, the transitional probability of *bi* given *bay* (*baby*) is computed as follows:

$$\frac{\text{frequency of } bay.bi}{\text{frequency of } bay}. \qquad [2]$$

The value computed in Eq. [2] is almost certainly greater than the transitional probability between the syllables in a word-external pair such as *bay#too:*

$$\frac{\text{frequency of } bay\#too}{\text{frequency of } bay}. \qquad [3]$$

While both Eqns. [2] and [3] contain the same denominator, the numerators will differ because word-internal pairs are generally more frequent than word-external pairs. This is likely to be particularly true with the relatively small lexicon found in infant-directed speech. Consequently, the transitional probabilities will tend to be lower for those syllable pairs which straddle word boundaries.

While the differences between word-internal and word-external transitional probabilities may be small, this information will be available in the input to learners. Moreover, this cue to word boundaries should be present in all natural languages. A learner, then, might hypothesize word boundaries upon discovering troughs in the transitional probabilities between syllables. Note, however, that this is a computational endeavor of considerable complexity. For this reason, distributional cues have not been widely studied by re-

searchers interested in word segmentation (although see Goodsitt et al., 1993). For the most part, recent research concerning distributional cues to word segmentation has been limited to computer models. One such model demonstrates that the minimum description length principle[2] can provide appropriate segmentations given distributional information (Brent & Cartwright, in press). Other models indicate that class-based n-gram and feature-based neural network models can segment speech using transitional probabilities (Cairns, Shillcock, Chater, & Levy, 1994); similarly, Elman (1990) describes a simple recurrent network able to discover written words in text by computing graded co-occurrence statistics (see Wolff, 1975, for similar findings using a non-connectionist architecture). These corpus-based models, along with many others in the machine speech recognition literature, demonstrate that statistical information is sufficient *in principle* for rudimentary word segmentation. Except for the Hayes and Clark (1970) study, however, data concerning human abilities to acquire such information are lacking.

One might argue that transitional probabilities are a needlessly complicated statistic for humans to use in solving this problem, and that simple co-occurrences of pairs of sounds would do just as well. In fact, simple co-occurrences are insufficient to solve this task, as they do not take into account the individual frequencies of each sound. To take an example, *the#dog* is likely to be a frequent pair of syllables. Were word boundaries determined based on simple frequency of co-occurrence, *the#dog* would be a perfectly reasonable lexical hypothesis for a single two-syllable word. Alternatively, if the high frequency of *the* is taken into account, as it would be by a mechanism computing transitional probabilities, then it becomes clear that the correlation between instances of *the* and *dog* is actually quite low.

The distinction between these different sta-

tistics may be seen in the form of children's errors. Consider the following oversegmentation error: the parent says ''Behave!'', and the child replies, ''I am/heyv/'' (Peters, 1985, p. 104). Here, the child, presumably aware of the overall frequency and grammatical function of the syllable *be,* has mistakenly segmented /heyv/ as an individual word. This error most likely reflects transitional probability and the semantics of the situation; otherwise, *behave* should be a perfectly good lexical item based on co-occurrence alone. The same is true for undersegmentation errors (common errors of this type include treating a phrase like ''ham'-neggs'' as a single word). Brown (1973) concludes a discussion of such errors by stating that for pairs which are mistakenly joined as single morphemes, ''the transition probability from the first member to the second has a relatively high value'' (p. 396).

Clearly, as mentioned earlier, transitional probabilities computed over syllables are just one of a family of possible statistical cues to word boundaries. For example, the transitional statistics of phoneme bigrams (or *n*-grams) are also potentially useful cues. Whether statistics are computed across phonemes, syllables, or other subword units (for example, moras in Japanese), the same types of learning mechanisms should suffice in principle for the induction of word boundaries. This is an important point, as many languages (including English) do not exhibit clear syllabification (Cutler, Mehler, Norris, & Segui, 1986). It has been argued that early infant speech perception is universally characterized by course-grained representations, roughly the size of syllables (Jusczyk, 1993; Mehler, Dupoux, & Segui, 1990), and that in particular, English-speaking 2- to 3-month-olds represent speech in syllabic rather than phonetic units (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Jusczyk & Derrah, 1987). Nonetheless, whatever the units in which early speech perception is framed, transitional probabilities computed over these units will serve to locate word boundaries.

The present research further investigates the hypothesis that humans are sensitive to transitional probabilities as cues to word bound-

---

[2] Intuitively, this algorithm provides an evaluation function which minimizes the amount of memory needed to represent a lexicon derived from a previously unsegmented corpus of speech.

aries. If learners can use this type of information, then the distribution of subword units in the input could serve as an initial bootstrapping device for generating candidate lexical hypotheses.

EXPERIMENT 1

We designed an artificial language learning task to investigate adult subjects' ability to use transitional probabilities as cues to word boundaries. Unlike the experiments in Hayes and Clark (1970), the language used here consisted of English phonemes (four consonants and three vowels, used to generate six trisyllabic words). Linguistic stimuli were used in order to make the task more natural; however, it was unknown prior to conducting the experiment whether English-like stimuli (in comparison with Hayes and Clark's warbles and glides) would make the subjects' task easier, because these stimuli are easier to represent and remember, or more difficult, due to interference from prior knowledge of English.

A second divergence from Hayes and Clark (1970) concerns the uniqueness relations in the stimuli. Each phoneme occurred twice on average in the lexicon used by Hayes and Clark. In the lexicon of the present language, however, each phoneme occurred approximately five times. Thus, each phoneme occurred more often and in more words on average in the present language than in Hayes and Clark's language, increasing the complexity of the task. The exposure duration was decreased to 21 min, less than half of the exposure duration used by Hayes and Clark (1970).

In the present experiment, learning was tested in one of two ways. Half of the subjects were given a forced-choice test between words from the language and nonwords, which consisted of syllables from the language in a novel order, unexemplified in the speech stream. This test was intended to determine whether subjects had learned anything about the order of syllables heard in the speech input. The other half of the subjects received a forced-choice test between words and part-words, which each consisted of two syllables from a word plus an additional syllable. This test provided a stronger measure of learning;

to succeed, subjects were required to discriminate words from part-words differing by only one syllable.

Method

Subjects. Twenty-four undergraduates were recruited from introductory Psychology classes at the University of Rochester. Subjects received course credit for their participation. All subjects were monolingual English speakers.

Materials. The language consisted of four consonants (p, t, b, d) and three vowels (a, i, u) which, when combined, rendered an inventory of 12 CV syllables. These were then combined to create six trisyllabic words: (*babupu, bupada, dutaba, patubi, pidabu,* and *tutibu*). Some members of the syllable inventory occur in more words than others; for example, *bu* occurs in four words, while *ta* occurs in only one. This was done to ensure varying transitional probabilities within the words themselves (range 0.31 to 1.0). The transitional probabilities between syllables spanning a word boundary were lower, as one would expect, and ranged between 0.1 and 0.2. These two ranges are likely to be even closer to one another in natural languages.

Three hundred tokens of each of the six words were concatenated into a text in random order, with the stipulation that the same word never occurred twice in a row. All word boundaries were removed from the text, rendering a list of 4536 syllables. The text was then read by the MacinTalk speech synthesizer, using the text-to-speech application Speaker, running on a Quadra 700 computer. As the word boundaries had been previously removed from the text, the synthesizer was unaware of them. Thus, the synthesizer did not insert any acoustic word boundary cues and produced equivalent levels of coarticulation between all syllables. The speech stream contained no pauses and was produced by a synthetic female voice in a monotone at 216 syllables per minute. The output of the synthesizer was tape-recorded directly from the sound output jack of the Quadra using a Sony Walkman Pro and broken into three listening blocks of seven minutes each.

For the test phase, six nonword foils and six part-word foils were created. The nonwords consisted of syllables from the language's syllable inventory which never followed each other in the speech stream, even across word boundaries. Thus, the transitional probabilities between each of the syllables in the nonwords were zero. Each of the six part-words consisted of a syllable pair from a word from the language, plus an additional syllable. For example, the final syllable of the word *pidabu* was altered to create the part-word *pidata.* Similarly, the first syllable of the word *dutaba* was altered to create the part-word *bitaba.* Three part-words contained the first two syllables of words, like *pidata,* and three contained the final two syllables of words, like *bitaba.*

The stimuli for the test phase were also generated by the MacinTalk speech synthesizer. Each of the words and foils was synthesized separately on the Quadra, recorded with the Sony Walkman Pro, and digitized onto a Macintosh IIsi computer using MacRecorder and SoundEdit Pro.

Apparatus. The experiment was conducted in a IAC sound-attenuated booth. The tape-recorded speech stream was presented to subjects using an Aiwa tape deck and a Proton speaker. The test phase was administered on a Macintosh IIsi computer, using the PsychLab stimulus presentation package.

Procedure. Subjects were instructed to listen to a ''nonsense'' language. They were told that the language contained words, but no meanings or grammar. They were informed that their task was to figure out where the words began and ended. Subjects were given no information about the length or structure of the words or how many words the language contained. They were informed that the listening phase of the experiment consisted of three short blocks, followed by a test of their knowledge of the words in the language. Subjects were given a 5-min break after each of the first two 7-min listening blocks. All subjects were run individually.

After a total of 21 min of listening, subjects received a two-alternative forced-choice test. For each test item, subjects heard two trisyllabic strings, separated by 500 ms of silence.

One of these strings was a word from the nonsense language, while the other was not. Subjects were asked to indicate which of the two strings sounded more like a word from the language by pressing either the ''1'' or ''2'' key on the keyboard. Four practice trials were given to each subject prior to the test, in order to clarify the structure of the test and to give subjects practice with the key presses. The practice trials consisted of strings of syllables which were not part of the language's syllable inventory; subjects were told that the practice trials did not have a correct answer. The test items were constructed by pairing the six words of the language with one of two types of trisyllabic foils. For half of the subjects ($n = 12$), the six nonword foils were illegal strings of syllables from the language. For the other half of the subjects ($n = 12$), the six foils were part-words, consisting of a syllable pair from a word plus an additional syllable. Each word was paired exhaustively with each foil in both tests, rendering 36 trials per test. There was a 2-s interval following the subject's response on each trial prior to the onset of the next trial. The order of presentation of items was randomized for each subject.

Results and Discussion

The overall results are presented in Fig. 1. In the nonword test condition, the mean score was 27.2 of a possible 36 (76%), where chance performance equals 18. A single-sample t test (two-tailed) showed overall performance significantly different from chance: $t(11) = 6.78$, $p < .01$. Each of the six words was learned at a level significantly better than would be expected by chance ($p < .01$ for each word). The performance by subjects on the part-word test was somewhat worse, as the foils in this test were more confusable with the words. The mean score was 22.3 of a possible 36 (65%), which was significantly worse than performance on the nonword test: $t(22) = 2.66$, $p < .05$. However, subjects in this more difficult condition still performed significantly better than would be expected by chance: $t(11) = 3.38$, $p < .01$. Three of the six words were learned significantly better than would be ex-
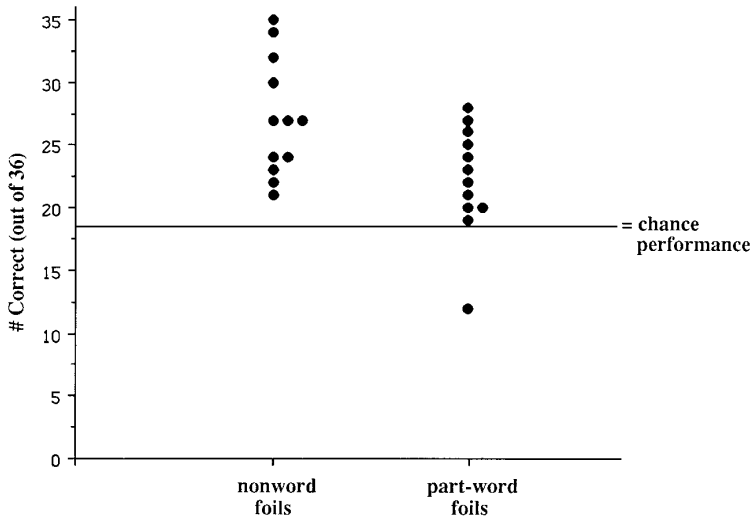
FIG. 1. Mean scores for each subject in the two conditions of Experiment 1, of a possible 36 correct.

pected by change ($p < .05$ for one word, and $p < .01$ for two words).

Next, we asked whether the strengths of the transitional probabilities played a role in which words were learned better than others. Recall that the transitional probabilities between the pairs of syllables within words ranged between .31 and 1. We split the six words into two sets, one set containing the three words with the highest average transitional probabilities (1.0, .75, and .75), and the other set containing the three words with the lowest average transitional probabilities (.5, .41, and .37).[3] A matched-pairs t test comparing subjects' mean performance on the three words with the highest transitional probabilities (79%) to the three words with the lowest transitional probabilities (72%) indicated that performance was superior on the words containing higher transitional probabilities: $t(11) = 2.31$, $p < .05$. These data support the hypothesis that the strength of the statistical relationship between pairs of syllables affects subjects' ability to learn word-like units.

To ensure that subject performance was not

mediated by the simpler measure of syllable co-occurrence, we examined the effects of the relative frequency of pairs of syllables. Due to constraints created by the small size of the language, all word-internal syllable pairs save one occurred an equal number of times in the listening phase (300 times). One syllable pair (*bupa*) occurred more often: 420 times. If subjects were positing units based on simple co-occurrence information, then subjects in the part-word condition should have maintained the pair *bupa* as an early lexical hypothesis and erroneously chosen the part-word containing *bupa* when given the part-word test. In fact, however, the part-word *bupabi* triggered the fewest false alarms of all: it was selected on only 26% of the trials in which it occurred.

These results may be probed further to determine which aspects of the words were learned. Recall that the part-word test used two types of part-words. Half of the part-words contained the first two syllables of words plus a final syllable, and the other half contained the final two syllables of words plus an initial syllable. Interestingly, subjects tended to false alarm (choose incorrectly) when part-words contained the final two syllables of words, but not when part-words contained the initial two syllables of words.

---

[3] These values were computed by averaging the two transitional probabilities associated with each word: the transitional probability from syllable one to syllable two and the transitional probability from syllable two to syllable three.

Subjects false-alarmed on the three part-words containing the first and second syllables of words significantly less often than would be expected by chance: subjects chose these three part-words on only 29% of the test trials in which they occurred, $t(11) = -61.3$, $p < .01$. However, when presented with part-words containing the second and third syllables of words, subjects performed at chance levels, choosing the part-word on 48% of the trials: $t(11) = -.39$, $p > .05$. A matched-pairs t-test comparing average numbers of false-alarms on the two sets of part-words revealed that subjects incorrectly chose the part-words consisting of the final syllables of words significantly more often than those consisting of the initial syllables of words: $t(11) = -3.25$, $p < .01$.

These data indicate that subjects confused words with part-words which resembled the ends of words more often than with part-words which resembled the beginnings of words. It appears that subjects were learning the ends of words first. Interestingly, several other lines of research have found that the ends of words are most salient to language learners, beginning with Slobin's (1973, p. 412) operating principle: ''pay attention to the ends of words.'' For example, Echols and Newport (1992) have found that children's early productions tend to include final syllables more frequently than initial or medial syllables. Similarly, Woodward and Aslin (1990) and Fernald and Mazzie (1991) found that mothers tend to place new words at the ends of utterances, prior to pauses, thereby highlighting the final syllable of the final word in the utterance.

Why would the ends of words be most salient to learners? Logically speaking, any cue which signals the end of a word also signals the beginning of another. If learners can find cues for word boundaries, why are these cues used to learn about the end of the last word rather than the beginning of the next word?

One possible account concerns the way that transitional probabilities are computed. The mechanism outlined in [1] computes, for each syllable, the probability that the next syllable will follow. A word boundary is hypothesized if the transitional probability is low. In order

to discover a word boundary between *da* and *pu,* for example, the learner must keep track of how often *da* occurs and how often *dapu* occurs:

transitional probability of

$$pu \,|\, da = \frac{\text{frequency of } dapu}{\text{frequency of } da} \quad [4]$$

If *dapu* occurs rarely relative to the overall frequency of *da,* then *dapu* is likely to cross a word boundary. Note that while this computation tells us both that *da* is the end of a word and that *pu* is the beginning of a word, the computation has been performed relative to the overall frequency of *da,* not *pu.* It is possible that the learner tends to maintain the information that *da* is the end of a word because *da* has anchored the computations required to discover a word boundary.

Note that this account of the end-of-word effect provides additional evidence for the existence of a mechanism which computes transitional probabilities to discover word boundaries. A predisposition to pay attention to the ends of words is a logical side effect of segmentation cued by transitional probabilities.

Apparently, then, distributional cues are not too complex for adult learners to exploit. Even when accompanied by no other cues, these rather complex statistical differences between word-internal and word-external sequences can be used by adult learners. This suggests that language learners may be sensitive to rather subtle aspects of the statistics of language input. Moreover, adult learners must be able to induce this information quite rapidly, given that they received only 21 min of exposure to the language.

Of course, natural languages contain a number of other cues which are at least probabilistically related to word boundaries, including prosody, semantics, inflectional morphology, and unstressed function words. The lack of other types of cues might have made the present task more difficult for the subjects, as only one type of cue was available. Alternatively, removing other types of information may have simplified the learning task, rendering a less

complex situation in which there were no competing sources of either relevant or irrelevant information. Experiment 2 addresses the effects of additional word-boundary cues on performance.

EXPERIMENT 2

In this experiment, distributional cues to words (transitional probabilities) were accompanied by a consistent prosodic cue, vowel lengthening. Final vowel lengthening serves as a cue to syntactic and word boundaries in a number of languages, including English (e.g., Klatt, 1975; Oller, 1973; Umeda, 1975). Furthermore, vowel length, along with pitch, amplitude, and vowel quality, cues syllabic stress cross-linguistically. Thus, vowel length is a prosodic cue that listeners are likely to be sensitive to and able to use in segmentation. Moreover, there exists ample evidence that young infants are sensitive to rhythmic structure in both linguistic and nonlinguistic tasks (e.g., Christophe et al., 1994; Morgan & Saffran, 1995; Trehub & Thorpe, 1989). It is therefore possible that vowel length is among the set of prosodic cues used by infants in early word segmentation.

For these reasons, Experiment 2 compared word segmentation for conditions in which vowel length accompanied distributional cues to a condition containing only distributional information. We chose to lengthen the vowels of only half of the words in each condition to enable comparisons of performance on prosodically cued versus nonprosodically cued words within subjects. In the initial lengthening condition, the vowels in the *first* syllables of three of the six words in the corpus were lengthened. In the final lengthening condition, the vowels in the *third* syllables of three of the six words were lengthened. The no lengthening condition, which served as the control condition, contained no prosodic cues and was therefore identical in structure to that in Experiment 1. Our predictions were as follows: if simply having a consistent prosodic cue correlated with a word boundary is sufficient to facilitate word segmentation, then learning in both of the prosodic conditions should be superior to the condition including only distribu-

tional cues. A second possibility is that third syllable (word-final) lengthening will facilitate performance due to the existence of word-final vowel lengthening in many languages.

In contrast to these predictions, Cutler, Mehler and their colleagues have suggested that speakers use their knowledge of the particular prosodic structure of their native language to segment unfamiliar speech (e.g., Cutler, 1990; Cutler, Mehler, Norris & Segui, 1992). In this case, subjects should use the prosodic structures which cue word boundaries in English to segment an artificial language. On this view, it is possible that performance will be superior in the initial lengthening condition, because vowel length is a correlate of stress in English and content words in English generally contain first syllable stress. This alternative corresponds to Cutler's (1990) Metrical Segmentation Strategy, a heuristic for the segmentation of English words from fluent speech whereby adult speakers assume that strong syllables correspond to the beginnings of words. Note, however, that only length cues, and not vowel quality and the other features which cue stress in English, are available to listeners in the present experiment.

Method

Subjects. Twenty-four undergraduates were recruited from introductory Psychology classes at the University of Rochester. Subjects received course credit for their participation. All subjects were monolingual English speakers.

Stimulus materials. The trisyllabic words were similar to those used in Experiment 1, except that the nasal consonants [m] and [n] were substituted for [b] and [d] respectively (*mupana, nutama, patumi, pinamu, timupu,* and *tumimu*). Lengthening was achieved by programming the MacinTalk synthesizer to lengthen specific vowels by 100 ms. In the initial lengthening condition, the first syllables in three of the six words were lengthened. In the final lengthening condition, the third syllables of three of the six words were lengthened. In the no lengthening condition, no prosodic cues were added, leaving transitional
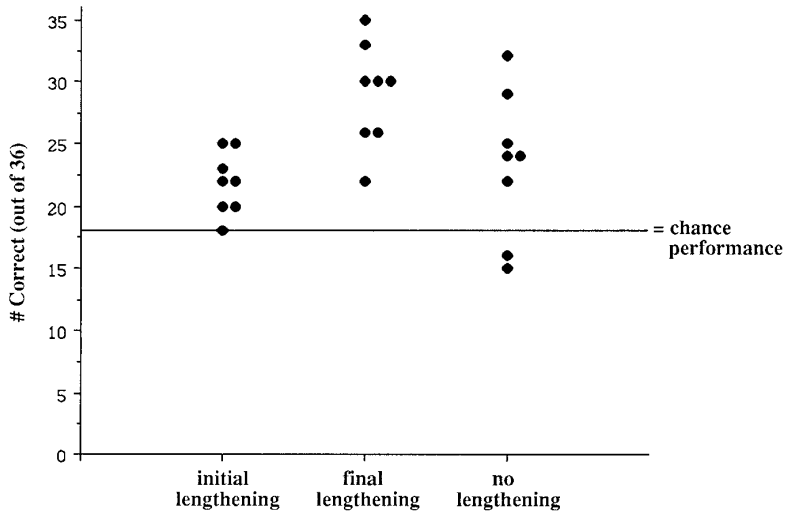
FIG. 2. Mean scores for each subject in the initial lengthening condition, the final lengthening condition, and the no lengthening condition.

probabilities as the only cue to word boundaries. The corresponding syllables of the nonword foils were also lengthened in the two lengthening conditions. In all cases, the syllable types lengthened in the words were the same as those lengthened in the nonwords; that is, if *mu* was lengthened as the first syllable of a word in the initial lengthening condition, then there would be a nonword beginning with a lengthened *mu.* This was done to ensure that subjects would not be able to succeed on the test simply by noting which syllable types were lengthened in the speech stream.

Apparatus. This was the same as in the previous experiment.

Procedure. This was the same as in the previous experiment, except that the between-subjects factor was the prosodic structure of the language. All subjects received the nonword foils (and not the part-word foils) in the forced-choice test.

Results and Discussion

Results from all three conditions are displayed in Fig. 2. Subjects in the initial lengthening condition performed with a mean score of 21.9 (61%); this mean was significantly different from chance: $t(7) = 4.4, p < .01$. The results for the no lengthening condition

replicate the first experiment, demonstrating that subjects can learn the words using distributional cues alone. These subjects ($n = 8$) performed with a mean score of 23.4 (65%), which was significantly different from chance: $t(7) = 2.6, p < .05$. The means of these two conditions are statistically equivalent: $t(14) = .672, p > .05$. However, subjects in the final lengthening condition performed significantly better than those in the other two conditions. The mean score was 29 (80%), which was significantly better than chance: $t(7) = 7.5, p < .01$. The *t* tests indicated that scores in the final lengthening condition were significantly better than both the initial lengthening condition ($t(14) = 2.23, p < .05$) and the no lengthening condition ($t(14) = 4.12, p < .01$).

In order to determine whether the lengthening cues facilitated the learning of individual words, we analyzed the data from each of the two prosodic conditions by comparing each subject's average score on the three words with lengthened syllables to their average score on the other three words. In the first syllable lengthening condition, subjects' mean scores on first-syllable lengthened words were not different from their scores on the other three words: $t(7) = 1.36, p > .05$. In the final syllable lengthening condition, however, the

comparison between mean scores on third-syllable lengthened words and the other three words was significant: $t(7) = 3.5$, $p = .01$. This pattern of data indicates that third-syllable lengthening paired with distributional cues is more effective than distributional cues alone in the segmentation of individual word units. First-syllable lengthening, on the other hand, is no more effective when paired with distributional cues than distributional cues alone.

The present results support the hypothesis that final syllable lengthening facilitates word learning, while initial syllable lengthening does not. These data suggest that subjects used their tacit knowledge of word-final lengthening to segment the speech stream. Such knowledge might stem from subjects' experience with the rhythmic structure of English. Another possibility is that word-final lengthening is a more frequently exemplified characteristic of languages. Research in acoustic phonetics has demonstrated that, in English and several other languages, the final syllables of words tend to be lengthened relative to the other syllables in the word (see, for example, Hoequist, 1983a, b, for data from English, Japanese, and Spanish speech).

While word-final lengthening is particularly prominent at syntactic boundaries, it is also present elsewhere in the utterance. Lengthening effects of this sort might perhaps reflect prearticulatory planning in preparation for the next unit of speech (Oller, 1973). According to Klatt (1976), features of speech which arise from production constraints may become the basis for perceptual rules and routines; were this the case with word-final lengthening, then one might expect that the detection of lengthening cues would become a part of the human perceptual arsenal of strategies for dealing with continuous speech. In any event, these results raise the possibility that word-final lengthening is a word-segmentation cue available to and usable by infants in many language environments.

Note, however, that even if such prosodic cues are available, and used by infants, distributional information is still required. It is not enough for the learner to notice that some syllables are lengthened. The learner *must also* determine whether there are regularities in this lengthening. Even if infants assume that lengthening will be correlated with word boundaries, English-learning infants must still learn that lengthening is correlated with the ends of words as opposed to the beginnings of words. In order to do so, learners must perform distributional analyses along with detecting prosodic cues (see Morgan (1996) for a related discussion of how limited distributional analyses might lead English-learning infants to adopt a trochaic bias). Lengthening, or any other prosodic cue, becomes informative as a predictor of word boundaries only when distributional cues are also taken into account. The learner may then use prosody and distributions in tandem to discover word units in the input.

GENERAL DISCUSSION

The present experiments demonstrate that adults are able to discover word units rapidly even in a system as impoverished as an unsegmented artificial language. This result is rather remarkable, given that such distributional cues have generally been considered to be too complex for human learners to use in language learning.

Of course, other cues are likely to coexist with distributional cues in natural language input. Prosodic cues are a particularly likely candidate for use by infant language-learners. This raises a difficult chicken and egg problem: does the infant use distributional analyses to discover prosodic regularities, or does she use prosody to isolate sections of speech upon which distributional analyses are conducted, along the lines of prosodic bootstrapping proposals? (See the collection of papers in Morgan & Demuth, 1996, for recent discussions of the prosodic bootstrapping hypothesis.) A recent series of studies by Morgan and Saffran (1995) provides some evidence for the latter account. Nine-month-old infants appeared initially to form units based on the rhythmic regularities in simple syllabic combinations and subsequently to analyze the distributional regularities. These data indicate that 9-month-old infants attempt to integrate prosodic and distributional cues when segmenting speech.

More generally, it seems quite likely that language learners take advantage of any and all cues which they can discover in the language input. Correlated sources of information, extracted by appropriately biased learning mechanisms, should eventually converge on solutions to learning problems like the discovery of word boundaries.

It remains to be seen how sensitive infants are to clusters of co-occurrences in somewhat more complex settings, as in the present artificial language studies. However, given that adult subjects are able to use such complex information, contrary to many assumptions about language learners, it seems likely that infants will also be able to exploit complex distributional cues. The ability to use probabilistic, distributional information would serve the child well not only in word segmentation, but also in solving many other difficult problems encountered in language acquisition.

Indeed, the fact that human learners are able to use distributional information in this fashion leads us to reconsider more broadly the role of statistical information in language learning. To take just one example, the acquisition of form class information might be facilitated by statistical analyses of both the distributional patterns of words (Finch & Chater, 1994; Maratsos & Chalkley, 1981; Mintz, Newport, & Bever, 1995) and the prosodic patterns of words (Kelly, 1992). On this view, distributional learning may be an accompaniment, or even perhaps a reasonable alternative, to certain assumptions concerning innate knowledge of language. Moreover, there are cases where distributional analyses are necessary components of language learning, including the acquisition of semantically arbitrary grammatical gender systems and other aspects of morphology and syntax.

One plausible objection to this line of argumentation is that while statistical mechanisms might accomplish such feats of learning, we have as yet very little direct evidence that humans actually make use of the statistics of the language input. To counter this objection, we would offer the host of data suggesting that adults maintain rich representations of far-flung statistical features of their language,

ranging from word-frequency effects to probabilistic prosodic expectancies to orthographic bigram probabilities to frequency-based contingency effects in parsing (Cutler & Norris, 1988; Gernsbacher, 1984; Kelly, 1988; MacDonald, Pearlmutter, & Seidenberg, 1994; Morton, 1969; Seidenberg, 1987; see Kelly & Martin, 1994, for a cogent review of this literature). Moreover, far less impressive neural mechanisms than our own can compute and make use of such information: the rat, for example, is sensitive not merely to the frequency of CS–US pairings, but rather to the conditional probability of the US given the CS (Rescorla, 1968), much like the subjects in the studies described in this paper. The point here is not to reduce language learning to the statistics of language (or to reduce humans to rats), but rather to recognize that those aspects of language learning which are governed by experience must also have an innate component capable of organizing and making sense of that experience. To the extent that the linguistic information experienced by the child may be characterized as statistical, it seems likely that the innate arsenal will include biased statistical learning mechanisms designed to extract the regularities of natural language inputs.

REFERENCES

BERTONCINI, J., BIJELJAC-BABIC, R., JUSCZYK, P. W., KENNEDY, L. J., & MEHLER, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General,* **117,** 21–33.

BEST, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing during the first year of life.* Dordrecht: Kluwer.

BLOOMFIELD, L. (1933). *Language.* New York: Holt.

BRENT, M. R. (Ed.). (in press). Special issue of Cognition: Computational approaches to language acquisition. *Cognition.*

BRENT, M. R., & CARTWRIGHT, T. A. (in press). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition.*

BROWN, R. (1973). *A first language.* Cambridge, MA: Harvard University Press.

CAIRNS, P., SHILLCOCK, R., CHATER, N., & LEVY, J. (1994). Lexical segmentation: The role of sequential

statistics in supervised and un-supervised models. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Earlbaum Associates.

CHRISTOPHE, A., DUPOUX, E., BERTONCINI, J., & MEHLER, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America,* **95,** 1570–1580.

COLE, R., & JAKIMIK, J. (1980). *A model of speech perception.* Hillsdale, NJ: Lawrence Earlbaum Associates.

CUTLER, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing.* Cambridge, MA: MIT Press.

CUTLER, A. (1994). Segmentation problems, rhythmic solutions. *Lingua,* **92,** 81–104.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language,* **25,** 385–400.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology,* **24,** 381–410.

CUTLER, A. & NORRIS, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* **14,** 113–121.

ECHOLS, C. H. (1993). A perceptually-based model of children's earliest productions. *Cognition,* **46,** 245–296.

ECHOLS, C. H., & NEWPORT, E. L. (1992). The role of stress and position in determining first words. *Language Acquisition,* **2,** 189–220.

ELMAN, J. L. (1990). Finding structure in time. *Cognitive Science,* **14,** 179–211.

FERNALD, A., & MAZZIE, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology,* **27,** 209–221.

FINCH, S. P., & CHATER, N. (1994). Distributional bootstrapping: From word class to proto-sentence. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Earlbaum Associates.

GERKEN, L. A., LANDAU, B., & REMEZ, R. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology,* **26,** 204–216.

GERNSBACHER, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General,* **113,** 256–281.

GLEITMAN, L. R. (1990). The structural sources of verb meanings. *Language Acquisition,* **1,** 3–55.

GLEITMAN, L. R., & GILLETTE, J. (1995). The role of syntax in verb learning. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language.* Cambridge, MA: Blackwell.

GLEITMAN, L. R., & WANNER, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner and L. R. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge, MA: Cambridge University Press.

GOLINKOFF, R. M., & ALIOTO, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language,* **22,** 703–726.

GOODSITT, J. V., MORGAN, J. L., & KUHL, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language,* **20,** 229–252.

HALLÉ, P., & DE BOYSSON-BARDIES, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development,* **17,** 119–129.

HARRIS, Z. S. (1955). From phoneme to morpheme. *Language,* **31,** 190–222.

HAYES, J. R., & CLARK, H. H. (1970). Experiments in the segmentation of an artificial speech analog. In J. R. Hayes (Ed.), *Cognition and the development of language.* New York: Wiley.

HOEQUIST, C. (1983a). Durational correlates of linguistic rhythm categories. *Phonetica,* **40,** 19–31.

HOEQUIST, C. (1983b). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica,* **40,** 203–237.

HYMAN, L. M. (1977). On the nature of linguistic stress. In L. M. Hyman (Ed.), *Studies in stress and accent.* Southern California Occasional Papers in Linguistics, No. 4. Los Angeles: University of Southern California.

JUSCZYK, P. W. (1993). How word recognition may evolve from infant speech perception capacities. In G. T. M. Altmann and R. Shillock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting.* Hillsdale, NJ: Lawrence Earlbaum Associates.

JUSCZYK, P. W., & ASLIN, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology.*

JUSCZYK, P. W., CUTLER, A., & REDANZ, L. (1993). Infants' sensitivity to predominant stress patterns in English. *Child Development,* **64,** 675–687.

JUSCZYK, P. W., & DERRAH, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology,* **23,** 648–654.

JUSCZYK, P. W., FRIEDERICI, A. D, WESSELS, J. M., SVENKERUD, V. Y., & JUSCZYK, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language,* **32,** 402–420.

JUSCZYK, P. W., LUCE, P. A., & CHARLES-LUCE, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language,* **33,** 630–645.

KELLY, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language,* **27,** 343–358.

KELLY, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical cat-

egory assignment. *Psychological Review,* **99,** 349–364.

KELLY, M. H., & MARTIN, S. (1994). Domain-general abilities applied to domain-specific tasks: Sensitivity to probabilities in perception, cognition, and language. *Lingua,* **92,** 105–140.

KLATT, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics,* **3,** 129–140.

KLATT, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America,* **59,** 1208–1221.

KUHL, P. K., WILLIAMS, K. A., LACERDA, F., STEVENS, K. N., LINDBLOM, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science,* **255,** 606–608.

LYONS, J. (1968). *Introduction to Theoretical Linguistics.* Cambridge: Cambridge University Press.

MACDONALD, M. C., PEARLMUTTER, N. J., & SEIDENBERG, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review,* **101,** 676–703.

MARATSOS, M., & CHALKLEY, M. (1981). The internal language of children's syntax. In K. E. Nelson (Ed.), *Children's language* (Vol. 2). New York: Gardner Press.

MEHLER, J., DUPOUX, E., & SEGUI, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing.* Hillsdale, NJ: Lawrence Earlbaum Associates.

MINTZ, T. H., NEWPORT, E. L., & BEVER, T. G. (1995). Distributional regularities of form class in speech to infants. In J. Beckman (Ed.), *Proceedings of the North East Linguistic Society, 25,* Vol. 2. Amherst, MA: GLSA.

MORGAN, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language,* **35,** in press.

MORGAN, J. L., & DEMUTH, K. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Hillsdale, NJ: Lawrence Earlbaum Associates.

MORGAN, J. L., & SAFFRAN, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development,* **66,** 911–936.

MORTON, J. (1969). Interaction of information in word recognition. *Psychological Review,* **76,** 165–178.

OLLER, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America,* **54,** 1235–1247.

OSGOOD, C. E., & SEBEOK, T. A. (1965). *Psycholinguistics: A survey of theory and research.* Bloomington, IN: Indiana University Press.

QUINE, W. V. (1960). *Word and object.* Cambridge, MA: MIT Press.

PETERS, A. M. (1985). Language segmentation: Operating principles for the perception and analysis of language. In Slobin, D. I. (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.

RESCORLA, R. A.. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology,* **66,** 1–5.

SEIDENBERG, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancies? In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading.* Hillsdale, NJ: Lawrence Earlbaum Associates.

SEJNOWSKI, T. J., & NOWLAN, S. J. (1994). A model of visual motion processing in area MT of primates. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences.* Cambridge, MA: MIT Press.

SLOBIN, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development.* New York: Holt, Rinehart, & Winston.

TREHUB, S. E. & THORPE, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure. *Canadian Journal of Psychology,* **43,** 217–229.

UMEDA, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America,* **58,** 434–445.

WERKER, J. & TEES, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* **7,** 49–63.

WOLFF, J. G. (1975). An algorithm for the segmentation of an artificial language analogue. *British Journal of Psychology,* **66,** 79–90.

WOODWARD, J. Z., & ASLIN, R. N. (1990). Segmentation cues in maternal speech to infants. Poster presented at the International Conference on Infancy Studies, Montreal, Canada.