Much of the discussion in the field of lan-
guage acquisition focuses on the nature—nur-
ditchotomy, in particular. But, on the "na-
" 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 the-
etorical perspectives acknowledge that a sub-
hedral fact of the acquisition problem must
involve learning from the input language.
For example, even on a strong nativist/universal
view, one part of this learning must entail or-
dering 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 investiga-
tors have begun to consider the learning prob-
lem from the point of view of the rich and
complicated statistical proper-
ties 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 sur-
prisingly 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 (Jus-
czyk, Cutler, & Redanz, 1993; Jusczyk, Fried-
erci, Wessels, Svenkerud, & Jusczyk, 1993).
These distinctions are quite fine-grained: 9-
month-olds prefer phonetic patterns which oc-
cur frequently in their native language over
patterns which are also phonotactically legal
but which are relatively infrequent (Jusczyk,
Luce, & Charles-Luce, 1994). These prefer-
ences presumably reflect the infant’s ability to
monitor the relative frequencies of the pro-
sodic 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 ex-
posed to different languages.

A somewhat more difficult learning prob-
lem is seen in the developmental decline of
infant sensitivity to nonnative language pho-
netic contrasts. Unlike younger infants, year-
old infants only discriminate the phonetic con-
trasts 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 pres-
ent in the native language input but which are
not contrastive. Such learning requires the
ability to determine which exemplars are mem-
bers of the same phonetic category and which
are not, via their distribution in the in-
put. Similarly, Kuhl and her colleagues have
demonstrated that the vowel categories of six-
month-old infants are organized around proto-
types which differ as a function of the vowel
structure of the native language (Kuhl, Wil-
liams, Lacerda, Stevens, & Lindblom, 1992).
Once again, the structuring of the infant’s
speech perception presumably reflects the dis-
tribution of exemplars in the speech input.

Another complex type of learning which
must in large part be input-driven is word seg-
mentation. Unlike the spacing provided in
written text, spoken words are rarely bounded
by pauses (Cole & Jakimik, 1980), so the in-
fant 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 in-
vestigates 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 pro-
posed in the literature. One hypothesis is that
the child avoids the word-segmentation prob-
lem 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 (Bloom-
field, 1933). This solution is potentially prob-
lematic 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 presen-
tation of content words in isolation (although
see Geerken, Landau, & Remez, 1990, for evi-
dence 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 hypoth-
esis. This solution to the word-segmentation
problem places the burden of segmentation on
the child’s speech, whereas the assignment of
speakers generating such utterances to input.
One would therefore expect speakers to use
new words in isolation whenever possible in
their infant-directed speech. In fact, many
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 occasion-
ally resulting in ungrammatical structures),
pointing to a role for post-utterance pauses in
word-segmentation (see Golinkoff & Alotio,
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 pause, may play a role in the
discovery of
Another possibility is that children learn word boundaries by detecting which sounds co-occur with entities in the environment (e.g., Os- good & 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 compli- cated (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 labor- atory without the benefit of any nonlinguistic information (Jusczyk & Aslin, 1995), and 11- month-olds are able to recognize familiar words without any visual referent (Halle & 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; Goodlett, Mor- gan, & Kuhl, 1993; Hayes & Clark, 1970). Words may be defined in distributional terms, such as the statistics of the word’s phonemic context; the more constrained, the higher the correlations between phonemes. The size of the successor- words is therefore possible that subjects Haynes and Clark (1970) proposed that initial 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 speech. However, in order to have learned language-specific prosodic patterns of this type, the learner must also have some knowledge of word bound- aires 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.

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

statistical information over a corpus of utter- ances 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 environ- ment: via motion: the spatial–temporal corre- lations between the different parts of the mov- ing 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 lin- guists 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 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- words also exhibit internal stability: if a sequence is picked up and moved to another position in the sentence, it will remain the same word. Words are generally uninterrup- tably: extraneous information, such as vowel length or stress, is generally placed be- tween words rather than in the middle of words. Words also exhibit internal stability: each lexical item consists of a sequence of phones in a fixed order—thus, when the se- quence is altered, a new word (or nonword) results.

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

As a result, they concatenated these sounds into four words and the words into a speech stream. Thus, 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 pho- nemes combinations found in the words, im- peded 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 sig- nificant, results, raising the possibility that learners’ abilities to acquire purely distribu- tional information might be limited. Subjects in their experiment scored only 62% correct on a two-alternative forced-choice test. More- over, 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, sepa- rated by pauses (e.g., aaa bbb ccc ddd), and the same four words with each pause shifted two phonemes to the left (aabb bbbb ccdd). Thus, on each of the 40 trials, one choice was always the same four words, though in differ- ent order. The second choice varied as a func- tion of the order of the words in the first choice, with a total of 24 possible sets of non- words. It is therefore possible that subjects might be achieved using a clustering mecha- nism measuring crude correlations between segments. Such a mechanism could identify a 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 synthesisizing a set of nonlinguistic sounds (glides and warbles) to represent “pho- nemes”, 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.
of the learning process based on their experi-
ence as pilot subjects: “after perhaps a minute of
listening, an event...stands out from the
stream...the listener may notice that it is typi-
cally preceded or followed by another event.
Recognition of a word, then, seems to proceed
from perceptually distinctive foci outwards in
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 sta-
tistical measure: transitional probability. A
simple example will illustrate the intuitions
underlying the relationship between transi-
tional probability and words in a natural lan-
guage. Note that while we present this exam-
ple 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 sublexi-
cal units. Consider the word baby. Bay is a
fairly frequent syllable in English, which will
be followed by in 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 tic, 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#not, bay#one, and bay#
ical, 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 transi-
tional probability. The transitional probability
of Y given X is computed as follows:

\[
\text{frequency of pair } XY \over \text{frequency of } X
\]

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 transi-
tional probability of bi given bay (baby) is
computed as follows:

\[
\text{frequency of bay.bi} \over \text{frequency of bay}
\]

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

\[
\text{frequency of bay#too} \over \text{frequency of bay}
\]

While both Eqs. [2] and [3] contain the same
denominator, the numerators will differ be-
cause 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 probabilisti-
cies 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 dis-
covering troughs in the transitional probabiliti-
ies between syllables. Note, however, that
this is a computational endeavor of consider-
able complexity. For this reason, distributional
cues have not been widely studied by re-
searchers interested in word segmentation (al-
though see Goodwin et al., 1993). For the most
part, recent research concerning distributional
cues to word segmentation has been limited
to computer models. One such model demon-
strates that the minimum description length
principle 1 can provide appropriate segmenta-
tions 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, Shill-
cock, Chater, & Levy, 1994); similarly, Elman
(1990) describes a simple recurrent network
able to discover written words in text by com-
puting 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
language recognition literature, demon-
strate that statistical information is sufficient
in principle for rudimentary word segmenta-
tion. Except for the Hayes and Clark (1970)
study, however, data concerning human abili-
ties to acquire such information are lacking.

One might argue that transitional probabilisti-
ces 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-occur-
rences are insufficient to solve this task, as
they do not take into account the individual
frequencies of each sound. To take an exam-
ple, 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 lexi-
cal hypothesis for a single two-syllable word.
Alternatively, if the high frequency of the is
taken into account, as it would be by a mecha-
nism computing transitional probabilities,
then it becomes clear that the correlation be-
tween instances of the and dog is actually
quite low.

The distinction between these different sta-
1Intuitively, this algorithm provides an evaluation function which minimizes the amount of memory needed to re-
portics may be seen in the form of children’s
errors. Consider the following oversegmen-
tation error: the parent says “Behave!” , and
the child replies, “I am#heyl!” (Peters, 1985, p.
104). Here, the child, presumably aware of the
overall frequency and grammatical function of
the syllable he, has mistakenly segmented
heyl 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-
eggs” as a single word). Brown (1973) con-
cludes 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 type of learning mecha-
nisms should suffice in principle for the induc-
tion of word boundaries. This is an important
point, as many languages (including English)
do not exhibit clear syllabification (Cutler,
1988; Jusczyk & Derrah, 1987). Nonetheless,
this conundrum 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 syl-
labic rather than phonetic units (Bertoncini,
Bijeljac-Babic, Jusczyk, Kennedy, & Mehler,
1996; Jusczyk & Derrah, 1987). Nonetheless,
whatever the units in which early speech per-
ception 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 transi-
tional probabilities as cues to word bound-
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’s (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 in 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: (babuqa, buapada, dutaba, patubi, pidaba, and tutiba). Some members of the syllable inventory occur in more words than others; for example, tuu 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 closer to one another in natural languages. Each 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. Each of the six 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 what the words were. 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 nonlanguage, while the other half of the 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 familiarize the subject with the words and foils. 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 nonwords 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%), which chance performance equals 18. A single-sample t test (two-tailed) showed overall performance significantly different from chance: t(11) = 2.68, 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 22.5 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 monolingual 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-
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, \( t(11) = -6.13, p < .01 \). 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: \( t(11) = -3.39, p > .05 \). 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; 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 \( du \) and \( pu \), for example, the learner must keep track of how often \( du \) occurs and how often \( dapu \) occurs: 

\[
\text{transitional probability of } \frac{\text{frequency of } dapu}{\text{frequency of } du}
\]

If \( dapu \) occurs rarely relative to the overall frequency of \( du \), then \( dapu \) is likely to cross a word boundary. Note that while this computation tells us both that \( du \) 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 \( du \), not \( pu \). It is possible that the learner tends to maintain the information that \( du \) 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 cues 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 simple 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 distributional 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. 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 probabilities as the only cue to word boundaries. The corresponding syllables of the non-word 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) = .72, *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, Hoegg, 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 syllable (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 choice for the learner: 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 based, along the lines of prosodic bootstrapping proposed? (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. This pattern of data indicates that third-syllable 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.

**References**


CAIRNS, P., SHELLFORD, R., & LEVY, J. (1994). Lexical segmentation: The role of sequential
WORD SEGMENTATION


Received February 23, 1995
Revision received June 28, 1995