Running Head:
DISCOVERING NOVEL WORD-LIKE UNITS FROM UTTERANCES

On the discovery of novel word-like units from utterances:

An artificial-language study with implications for native-language acquisition

Delphine Dahan and Michael R. Brent

Johns Hopkins University

Abstract

In 4 experiments, adults were familiarized with utterances from an artificial language. Short utterances occurred both in isolation and as part of a longer utterance, either at the edge or in the middle of the longer utterance. After familiarization, subjects' recognition memory for fragments of the long utterance was tested. Recognition was greatest for the remainder of the longer utterance after extraction of the short utterance, but only when the short utterance was located at the edge of the long utterance. These results support the INCDROP model of speech segmentation and word discovery, which asserts that people segment utterances into familiar and new word-like units in such a way as to minimize the burden of processing new units. INCDROP suggests that segmentation and word discovery during native language acquisition may be driven by recognition of familiar units from the start, with no need for transient bootstrapping mechanisms.

Children typically learn their first words at about ten or twelve months of age. They learn words slowly for the next six or eight months, averaging roughly one word per day. At about eighteen to twenty months, they begin to use two-word combinations much more frequently and their rate of vocabulary acquisition increases markedly. To acquire each new vocabulary item, a child must associate the sound pattern of a word with its meaning and, at some point, its syntactic function. Young children's ability to discover and memorize the meaning and syntactic function of each sound pattern would be remarkable enough to warrant serious scientific investigation even if the sound patterns themselves were obvious from the linguistic environment. However, speech does not appear to contain any reliable acoustic marking of word boundaries, so the input children receive is better modeled as a sequence of unsegmented utterances than as a sequence of words — the sound patterns of individual words must be discovered.

This paper focuses on processes by which children could identify word-like units to use as candidate sound patterns for association with meaning. The investigation is motivated by the INCDROP model of speech segmentation and word discovery (Brent, 1997; see also Brent, 1996; Brent & Cartwright, 1996). INCDROP asserts that the process of segmenting utterances and inferring new word-like units is driven by the recognition of familiar units within an utterance. By utterance, we mean a complete act of speaking, surrounded by silent pauses and ending in the prosodic characteristics of clause boundaries. In particular, INCDROP predicts that familiar units will tend to be extracted from an utterance, and the remaining contiguous stretches will be inferred as novel units. For example, if look is recognized as a familiar unit in the utterance Lookhere! then look will tend to be segmented out and the remaining contiguous stretch, here, will be inferred as a new unit. As a special case of this general principle, INCDROP predicts that utterances containing no familiar units will be treated as a single novel unit and stored in memory. Thus, no special “bootstrapping” or initialization mechanism is necessary to start the process of discovering new units by extracting familiar units. Our working hypothesis is that the INCDROP segmentation and word-discovery mechanisms are available throughout life, from the first word learned to the last. In the present study, we set out to test some of INCDROP’s predictions with a series of experiments in which adults listened to miniature artificial languages.

Bootstrapping the lexicon

A number of proposals have been put forth to explain how children discover their first word-like units. These proposals are motivated by the notion that children cannot begin to use a segmentation strategy based on the extraction of familiar words until they already know some words. Thus, it has been proposed that young children use certain transient strategies to discover their first words. Eventually, as children learn more words, the recognition of familiar words comes to play a greater role and the early strategies recede into the background, perhaps continuing to play a role as cues to speed the recognition of familiar words during on-line processing. Two classes of proposed strategies for discovering the first words are those based on stress and those based on transitional probability.

Based on Slobin’s (1973) operating principle “Pay attention to salient parts of speech,” Gleitman and her colleagues suggested that stressed syllables (in stress-accent languages like English) might help young children identify their first words (Gleitman, Landau, & Wanner, 1988; Gleitman & Wanner, 1982). Echols and Newport (1992) reported that the first words produced by English-learning children tend to omit unstressed syllables and syllables that are not word-final. This suggests that perceptual and attentional biases for stressed syllables and word-final syllables may assist in the initial extraction of word-like units. Jusczyk, Cutler, and colleagues have proposed that infants use stressed syllables to hypothesize the beginning of words in fluent speech (Cutler, 1994, 1996; Houston, Jusczyk, & Newsome, 1995; Newsome & Jusczyk, 1995). This proposal is related to the Metrical Segmentation Strategy (Cutler, 1990; Cutler & Carter, 1987; Cutler & Norris, 1988), which was first put forth as a strategy used by English-speaking adults to optimize lexical access by hypothesizing word onsets at strong syllables.[1] Jusczyk and his colleagues (Jusczyk, 1997; Jusczyk & Aslin, 1995; Houston, Jusczyk, & Newsome, 1995; Newsome & Jusczyk, 1995) found evidence that 7 ½-month-old infants can segment words beginning with a strong syllable, like DOCtor, out of fluent speech but not words beginning with an unstressed syllable, like guiTAR (see also Echols, Crowhurst, & Childers, 1997; Morgan, 1996). Further, they found that if the stressed syllable was consistently followed by the same syllable, infants would treat the two-syllable sequence as a single word-like unit; if the context after the stressed syllable varied, however, infants seemed to restrict the extracted unit to the stressed syllable. This suggests that 7 ½-month-old infants are sensitive to distributional regularities as well as preferring stressed-syllable onsets when constructing word-like units. Unlike 7 ½-month-olds, 10 ½-month-olds were able to extract units that started with an unstressed syllable, suggesting that word segmentation is less constrained by prosodic factors at this age. Morgan and his colleagues (Morgan, 1994; Morgan & Saffran, 1995) also found that rhythm (the alternation of stressed and unstressed syllables) and regularities of syllabic sequence led infants to cluster syllables as a unit; moreover, they found that 9-month-old infants were sensitive to the consistency between rhythmical and distributional factors, whereas 6-month-olds grouped syllables by rhythmic regularity, regardless of whether syllable-sequence regularity was also present.

Transitional probabilities form the basis for another class of proposals about how children segment utterances into word-like units before they know any such units (Goodsitt, Morgan, & Kuhl, 1993; Hayes & Clark, 1970; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). The transitional probability between two phonemes or syllables *x* and *y* is defined to be the proportion of occurrences of *x* that are followed by *y*. Transitional probabilities have been argued to be higher within words than between words, since the pairs of adjacent phonemes or syllables within words are constrained by the lexicon and by phonotactics, whereas pairs of adjacent phonemes or syllables spanning a word boundary are less constrained. Thus, it has been proposed that children posit word boundaries at points of low transitional probability. Saffran, Aslin, and Newport (1996) showed that after exposure to a continuous stream of syllables, 8-month-old infants showed differential durations of listening for sequences of syllables that had always occurred successively and sequences that had occurred successively less often, suggesting a sensitivity to statistical relationships between neighboring speech sounds. According to this segmentation strategy, word-boundary locations at any given time are based on the transitional-probability computations of the utterances heard before. However, transitional-probability computations do not take into account the segmentation points in previous utterances; in other words, having isolated some words does not help isolating other words or even the same words later on.

The proposal that young children discover new words based on transitional probabilities is motivated as a temporary strategy that infants might use to extract their first word-like units. Saffran, Newport, and Aslin (1996) suggest that its role would be limited to serving "as an initial bootstrapping device for generating candidate lexical hypotheses" (p. 611).

According to the INCDROP model, however, there is no need to assume a special strategy for segmenting the very first words. The same word-segmentation principles are applied at all stages of language acquisition (and possibly in lexical access for adults, see Brent, 1997). Children segment each utterance by recognizing and extracting units they have already discovered. If an utterance does not contain any familiar units, or if no units at all have been discovered yet (the “empty lexicon” case), the utterance is treated as a single unit and stored in memory. This default behavior provides some familiar units that can then be used to segment later utterances. For example, if a child heard Getit! she or he would treat the entire utterance as a single novel unit, erroneously in this case. If he or she later heard yougetit? then the getit part would be recognized as familiar and segmented out, isolating the novel unit you. Since child-directed speech contains many such short utterances (e.g., Aslin, Woodward, LaMendola, & Bever 1996; Bernstein Ratner, 1996; Brown & Bellugi, 1964; Newport, Gleitman, & Gleitman, 1977; Snow, 1972), it is not difficult to find such examples to prime the segmentation and word-discovery pump. Thus, INCDROP differs from other proposals in asserting that the child’s experience with the word-like units of the language, rather than the statistical relationships between phonological units, is the primary determinant of early segmentations. Since the segmentation of an utterance is directly related to the lexical units stored in memory, INCDROP can be described as a model in which segmentation is lexically driven.

It is important to stress that INCDROP does not exclude the possibility that other segmentation strategies play a role in the discovery of word-like units (see Christiansen, Allen, & Seidenberg, in press, for a multicue approach). However, our proposal does remove one theoretical motivation for such strategies — the apparent need for special priming mechanisms before new words can be isolated by recognizing familiar words.

The INCDROP model and its predictions

INCDROP, an acronym for INCremental Distributional Regularity OPtimization, is based on a mathematical model devised by Brent and Cartwright (1996; see also Brent, in press). In addition to their formulas, Brent and Cartwright presented a more abstract, qualitative characterization of the model, which was subsequently refined by Brent (1996, 1997). INCDROP is aimed at predicting how people segment a given utterance or sequence of utterances — that is, predicting which segmentation they will choose. The model assumes that each part of an utterance is either attributed to a familiar unit or attributed to a novel unit to be stored in memory. It attempts to predict which familiar and novel units each part will be attributed to. Brent (1997) presented three segmentation criteria stating that people segment an utterance in such a way as to: (a) minimize the total length of all novel words — that is, the fraction of the utterance attributed to novel words (and hence maximize the portion of the utterance attributed to familiar words); (b) minimize the total number of novel words; and (c) maximize the product of relative frequencies of all words. The relative frequency of a word is the number of times it has been encountered as a proportion of the total number of word tokens that have been encountered. The product of relative frequencies of a segmentation is the result of multiplying together the relative frequencies of the individual words in the segmentation, including novel as well as familiar words. Criteria (b) and (c) are closely related because novel words have the lowest possible relative frequency.[2] If one competing segmentation is favored over the alternatives because it implies fewer novel words, that same segmentation is favored because it has a greater product of relative frequencies. Since the number of novel words in a segmentation is a more intuitive quantity than the product of relative frequencies, we will focus on the number of novel words when both criteria support the same segmentation.

From a cognitive perspective, INCDROP can be interpreted as a “least effort” model in which segmentation is driven by an attempt to minimize the processing burdens of memorizing new words and accessing the memories of familiar words. The burden of memorizing new words increases with both the number and length of new words to be memorized. The burden of accessing the memories of familiar words decreases as the relative frequencies of the words to be accessed increase.

The INCDROP criteria can sometimes conflict, one criterion supporting one segmentation while another criterion supports a different segmentation. In that case, the predicted segmentation depends on how the total length of novel words, number of novel words, and product of relative frequencies are combined into a single number. In the mathematical model of Brent and Cartwright (1996), the logarithm of the product of relative frequencies is added to the other terms. In general, this suggests that the product of relative frequencies will tend to be outweighed by the other terms when they conflict. However, the precise resolution of conflicts depends on how raw length and frequency translate into cognitive burden, something about which we have little evidence at present.

Specific predictions

Although some of the general predictions of INCDROP have already been mentioned, it is worthwhile to see how the segmentation criteria lead to these predictions. Predictions for four cases are considered below: An utterance that contains no familiar word-like units, an utterance that contains one familiar unit at the beginning or end, an utterance that contains one familiar unit in the middle, and an utterance that contains two familiar units that overlap and hence compete with one another.

If an utterance does not contain any familiar word-like units, INCDROP’s segmentation criteria predict that it will be treated as a single novel unit, and will be stored in memory; thereafter, it is considered a familiar unit. Treating the whole utterance as one novel unit satisfies the segmentation criteria better than dividing it into multiple novel units. The single-unit analysis minimizes the number of novel units (and maximizes the product of relative frequencies), while the total length of novel units is the same regardless of how the utterance is divided up, since the utterance contains no familiar units.

If part of an utterance matches a familiar unit, the segmentation criteria predict that, in most cases, this familiar unit will be extracted and the remaining contiguous stretches treated as novel units. This is because extracting the familiar units reduces the total length of the novel units. However, if extracting a familiar unit implies the existence of novel units in the utterance then there is a conflict among the segmentation criteria. To get a better understanding of this conflict, consider the special case where the familiar unit is at the edge of an utterance that contains no other familiar units — for example, look is the only familiar unit in the utterance Lookhere! The competing segmentations include (a) extracting the familiar unit (look) and storing the remainder (here) as a novel unit, and (b) treating the entire utterance as one, long novel unit. Both options imply one novel unit, but the portion of the utterance attributed to novel words after extracting the familiar unit is smaller. Thus, the number of novel units is neutral, and the length of novel units favors extraction of the familiar unit. However, extracting reduces the product of relative frequencies by a factor of the relative frequency of the familiar unit.[3] The balance in this conflict depends on the length and frequency of the familiar unit: the longer it is, the better extraction satisfies the length criterion, and the more frequent it is, the less extraction violates the product of relative-frequencies criterion; units that are both short and rare may not be segmented out. Note that this is consistent with the observation that, across many languages, short words tend to have high frequency (Zipf, 1949). As mentioned above, however, the product of relative frequencies is weakened by a logarithmic transformation in the formulas of Brent and Cartwright (1996), suggesting that we should predict extraction of the familiar unit as the default, except in extremes of length and relative frequency.

If a familiar unit is in the middle of an utterance that contains no other familiar units — for example, if look is the only familiar unit in the utterance Don’tlookhere! — the situation is quite different. Extracting the familiar unit now implies two novel units, one on each side, while treating the entire utterance as a single, long novel unit implies only one. Thus, the number of novel units criterion changes from being neutral to opposing extraction. The opposition of the product of relative frequencies criterion is likewise strengthened, since novel units have the lowest possible relative frequency. If the familiar unit is sufficiently long and sufficiently frequent it should still be segmented out, but the model predicts that the extraction of familiar units from the middle of long utterances will impose a greater cognitive burden and hence will be rarer.

Finally, consider the case where more than one familiar unit competes for extraction, because the same portion of the input could be attributed to one unit or the other but not both. For example, in the utterance anicedog either an or nice can be extracted but not both, since the n cannot be attributed to both words. In such cases, the prediction depends on which segmentation of the entire utterance satisfies the segmentation criteria better. However, there is a special case where it is obvious that extracting one familiar unit satisfies the criteria better than extracting the other. When two familiar units of roughly equal length and frequency are the only familiar units in a longer utterance, and when one of the two is at the edge while the other is in the middle, then the one at the edge will be extracted in preference to the one in the middle. As discussed above, this is because extracting from the middle of the utterance implies two novel words, while extracting from the edge implies only one.

The role of the utterance in segmentation and word discovery

An important aspect of INCDROP is the central role accorded to the utterance, which serves three distinct functions. First, the utterance is a default unit, in the case where it contains no familiar units. Second, it provides the domain within which alternative segmentations are evaluated according to the model's criteria. That is, the model allows the locations of word boundaries at the beginning of an utterance to be influenced by sounds that occur at the end of the utterance, but such influences cannot extend across utterance boundaries. Third, utterance boundaries provide unambiguous word boundaries that are evident on the surface. Peters (1983) suggested that children might begin by storing the whole utterance as a single unit in the lexicon. Aside from Peters' approach, other approaches to segmentation and word discovery accord the utterance only the last of the three functions — supplying unambiguous word boundaries — and even this function is often left implicit. Thus, the central role accorded to the utterance within INCDROP is one of the most important differences between INCDROP and most other approaches.

Empirical evidence from a number of sources is consistent with the central role that INCDROP assigns to the utterance. First, infants as young as two months of age appear to treat the utterance as a single coherent unit (Mandel, Jusczyk, & Kemler Nelson, 1994; Mandel, Kemler Nelson, & Jusczyk, 1996). Turning to the role of the utterance as default unit, this strategy would be most effective if children were addressed with at least a few short utterances — ideally, though not necessarily, single-word utterances. Aslin, Woodward, LaMendola, and Bever (1996) reported that 17 out of the 19 mothers who they asked to teach novel words to their

12-month-olds produced at least some instances of the novel words in isolation, and that the proportion of utterances in which the novel word was presented in isolation reached 70% for some mothers. Although INCDROP predicts that isolated words are useful for getting the segmentation process started, it does not predict that they are necessary — short multiword utterances that can occur inside other utterances suffice, as illustrated by the Getit?...Yougetit! example, where the word you is discovered from two multiword utterances. The vast literature on child-directed speech uniformly reports the predominance of such short utterances in speech to children. The effectiveness of the INCDROP strategy receives further support from computer simulations reported by Brent and Cartwright (1996), in which INCDROP's segmentation criteria produced relatively good segmentations when applied to broad phonetic transcripts of child-directed speech. Finally, one of the most straightforward consequences of INCDROP is that an utterance that does not contain any familiar units is stored as a single unit, even though it may in fact contain more than one lexical unit. Several studies have reported the presence in children's early productions of formulas or formulaic expressions, apparent units that consist of more than one adult word (Hickey, 1993; Peters, 1983; Plunkett, 1993). The same phenomenon has been found for second-language learners (Hakuta, 1976; Vihman, 1982; Weinert, 1995; Wong Fillmore, 1976).

Limits of the INCDROP model

Despite this evidence supporting its plausibility, the INCDROP model remains an idealization that does not attempt to account for all the factors that might bear on segmentation and word discovery. In particular, the recognition of a familiar unit requires a certain degree of similarity between the representation of the familiar unit in memory and its acoustic realization in the context of a larger utterance. Because of coarticulation and other phonetic phenomena in continuous speech, this similarity may be low and recognition may fail, independently of the cognitive burden accounted for by INCDROP's criteria.

Moreover, INCDROP models the relative cognitive load involved in memorizing new units and accessing the memories of familiar units, but it does not attempt to model the limits of cognitive load at which this process fails. For utterances containing no familiar units, the model predicts that committing them to memory as a single, novel unit, will impose a smaller processing burden than other analyses of the same utterance. But that burden may still be very large, in absolute terms, if the utterance is long. In such cases, children may simply let the utterance go by without committing any new units to memory. Or, they may use some other segmentation strategy, such as a strategy based on stress or transitional probabilities, to break the long stretches up and commit only some of the smaller units to memory, abandoning others. Rather than add some arbitrary threshold of cognitive load into the model, we simply acknowledge that it will break down when even the best analysis is too difficult.

In addition to those cases where the cognitive burden of exhaustive analysis is too great, there are cases where multiple competing segmentations impose fairly similar burdens. In those cases, we would expect the choice among competing segmentations to be influenced by language-specific knowledge of other sorts, to the extent that the child has acquired such knowledge. For example, English-learning children appear to demonstrate a preference for words that begin with strong syllables from a very early age, perhaps reflecting the fact that the majority of English content-words begin with a strong syllable (Jusczyk, Cutler, & Redanz, 1993). If there are two competing segmentations of an utterance that are fairly equal in terms of the cognitive burden modeled by INCDROP, but one of them posits a novel word beginning with a weak syllable and the other posits a novel word beginning with a strong syllable, an English-learning child might well select the latter segmentation. Indeed, as the child forms more phonological generalizations about the way words typically sound in her language, these generalizations may come to play an increasing role in segmentation. For example, Brent and Cartwright (1996) discuss how children could learn phonotactic constraints on the consonant clusters that can occur at the beginnings and ends of words, and show via computer simulation that such constraints can greatly aid in segmentation and word discovery. Various kinds of probabilistic phonotactics, including transitional probabilities, could also play a role in characterizing what words typically sound like in the language being learned (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997).

The relationship between segmentation and chunking

Segmentation can be thought as the result of recoding long stimuli into smaller units or chunks. Chunking is a natural, perhaps automatic, tendency to process stimuli by parts (Bower & Springston, 1970; Johnson, 1970; Miller, 1956; Tulving, 1962). It has been observed on verbal materials, but in other cognitive domains as well, such as music perception (e.g., Deutsch, 1980; Dowling, 1973) and artificial-grammar learning using strings of letters (e.g., Servan-Schreiber & Anderson, 1990), which suggests that the mechanism for chunking rests on a general cognitive ability. The presence of chunks has been shown to directly influence the storage of stimuli in memory (Dowling, 1973), as well as the processing of subsequent stimuli: The more a subsequently presented novel string of letters can be divided into familiar chunks, the more familiar it seems (Buchner, 1994; Servan-Schreiber & Anderson, 1990; see Perruchet & Pacteau, 1990, where chunks are defined as bigrams of letters). Chunks can delimit low-level units, such as words, but can also be combined together and hence mark higher-level units, such as phrases and sentences (Servan-Schreiber & Anderson, 1990). Morgan, Meier, and Newport (1987) and

Valian and Levitt (1996) found evidence that prosodic phrasing, in highlighting structural relationships between words, can help learning the syntax of an artificial language. Chunks can be induced by the presence of blanks between groups of letters (Servan-Schreiber & Anderson, 1990), or by rhythmic cues, such as the lengthening of the last tone of a group and a pause between groups of tones (Dowling, 1973). In the present research, we aimed to show that chunking can be induced by previously stored chunks or familiar units, in the absence of any explicit boundary between units.

The present artificial-language study

We have cited evidence that infants recognize the coherence of utterances, evidence that the INCDROP strategy for segmentation and word discovery might be effective, and evidence that children sometimes treat a string of words as a single unit. However, there is no direct evidence that people segment and infer new word-like units by making use of the familiar units they have previously stored. The series of experiments presented here investigated the inference of new units by adults exposed to utterances from an auditory artificial language. Of course, evidence from adults learning an artificial language would not imply similar behavior by young children acquiring their native language. On the other hand, if we do find evidence that adults behave according to the predictions of INCDROP, there is no particular reason to believe that children would be different. Previous language-learning studies have shown similar findings for adults exposed to an artificial language and children or even infants exposed to the same type of language (Braine, Brody, Brooks, Sudhalter, Ross, Catalano, & Fish, 1990; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). This suggests that certain mechanisms involved in language learning are available to humans regardless of age (at least after very early infancy).

This study consisted of presenting adult listeners with utterances from a miniature artificial language, that is, sequences of nonsense syllables such as /dobuneripo/. The use of the artificial-language methodology enables one to experimentally manipulate the language-acquisition process with perfect control of the input (e.g., Brooks, Braine, Catalano, Brody, & Sudhalter 1993; MacWhinney, 1983; Meier & Bower, 1986; Moeser & Bregman, 1972; Valian & Coulson, 1988; Valian & Levitt, 1996). Our stimuli were designed to render the subjects' knowledge of their native language irrelevant to the task by including no English words, no syntactic or semantic information, and no consonant clusters. Further, the experimental design controlled for effects of the predominant stress patterns of English. The use of artificial language also allowed us to manipulate the factors that INCDROP predicts will influence segmentation and word discovery, such as the familiarity or novelty of possible units and their relative frequencies.

The nonsense utterances we presented to subjects were of two kinds: short utterances of two or three syllables, and long utterances of five syllables that contained the short utterance. We hypothesized that subjects would treat the short utterance in isolation as one unit, and would recognize the short utterance in the long utterance, segment it out, and store the remaining contiguous stretches of the long utterance as new unit(s). After exposure to the artificial-language utterances, we examined whether new units had been inferred, using two different paradigms. The first paradigm (used in Experiment 1) consisted of presenting subjects with test items and asking them to decide, for each one, whether it consisted of a word of the language they had just heard. Half of the test items corresponded to the predicted inferred unit after extraction of the familiar unit, while the other half consisted of smaller or larger fragments of the long utterance. We hypothesized that the test item corresponding to the inferred unit would feel more familiar to subjects. We expected the greater familiarity of the inferred units, as compared to longer or shorter fragments of the utterances, to be reflected in the explicit judgments made by subjects. Comparing performance on the inferred units with performance on items that did not respect the predicted inferred-unit boundaries directly addressed the segmentation and discovery of novel word-like units.

The second paradigm (used in Experiments 2, 3, and 4) consisted of a recognition-memory task. Subjects were presented with sequences of syllables and asked to decide, for each sequence, whether they had heard it during the familiarization phase. Among those test items that consisted of sequences of syllables that had occurred during familiarization, some of them exactly corresponded to the predicted inferred unit, while the others were shorter or longer. We hypothesized that the inferred units have their own representation in memory, whereas utterance fragments that are not inferred units can be accessed only by retrieving and analyzing the representation of the whole utterance. This suggested to us that subjects might be most accurate and fastest at remembering that they had heard a test item before if this test item corresponded to an inferred unit. Several findings in the literature support such a prediction. First, Johnston, Dark, and Jacoby (1985) showed that words are identified more quickly and accurately when presented for the second time than words that appear for the first time. Second, Dowling (1973) showed that subjects were better at remembering having heard a string of tone in the preceding longer string when the test string consisted of a chunk or unit induced by rhythm in the long string than when the test string straddled two rhythmic units in the long string. The same phenomenon was predicted to take place in the present study.

Experiment 1

The aim of Experiment 1 was to see whether subjects would extract a familiar word-like unit located at the edge of an utterance containing no other familiar units, infer that the remainder of the utterance is a novel unit, and store the novel unit in memory. We also wanted to see if it matters whether the familiar unit is located at the beginning or end of the utterance. Although INCDROP does not predict a difference, one can imagine that it might be easier to extract the familiar unit first, before hearing the entire utterance, than to hold the entire utterance in memory until the familiar unit is recognized at the end, then go backwards to construct a novel unit from the beginning of the utterance.

Before going into methodological details, it will be useful to outline the design principles and introduce some terminology. The experiment consisted of a familiarization phase and a test phase. The materials can be grouped into item sets, each consisting of several utterances presented during familiarization together with the corresponding test items. The familiarization utterances in each item set consisted of a two- or three-syllable short utterance (e.g., abc or ij, where each letter stands for a syllable) and a five-syllable long utterance which contained the short utterance either at the beginning or at the end. For example, the long utterance abcde begins with the short utterance abc, while the long utterance fghij ends with the short utterance ij. Each item set also contained one two-syllable test item and one three-syllable test item, each consisting of a fragment of the long utterance taken from the edge that did not contain the short utterance (e.g., de and cde from abcde, when abc is the short utterance). For the familiarization phase, all subject groups received the same long utterances, but they received short utterances of different lengths (e.g., abc or ab). The length of the short utterance was predicted to determine how the long utterance would be segmented, and thus what new unit would be inferred. For example, the presentation of the short utterance ab in association with the long utterance abcde would trigger the inference of the new unit cde; conversely, the presentation of the short utterance abc in association with the long utterance abcde would trigger the inference of the new unit de. For the test phase, each test item was said to be in the word condition when it consisted of exactly the fragment that remained after the short utterance presented during familiarization was extracted from the long utterance (for example, if ab was the short utterance and abcde the long utterance, cde was said to be in the word condition). Otherwise, it was said to be in the nonword condition. The design consisted of testing subjects who heard different familiarization phases on the same test items (see Redington & Chater, 1996, for methodological discussion). Each test item was tested an equal number of times in the word and nonword conditions. After the familiarization phase, subjects were presented with test items and asked to decide whether each test item was a word of the made-up language they had just heard. We predicted that subjects would judge a test item to be a word more often in the word condition (i.e., when it corresponds to a predicted stored unit) than in the nonword condition. Note that a yes response does not mean that a new unit has been isolated and inferred, nor does a no response mean that the inference has failed. Rather, a higher percentage of yes responses in the word condition would indicate greater subjective familiarity with these items, compared to the nonword items.

Method

Subjects

Forty-eight student volunteers from Johns Hopkins University were paid five dollars each for their participation. All were native speakers of English. They were tested separately in a sound-attenuated booth.

Materials

The materials consisted of four different item sets constructed out of consonant-vowel (CV) syllables that obeyed the minimal-word constraints on English CV words but were not actual words. The sets are presented in Table 1.

Insert Table 1 about here.

For each item set, the location of the short utterance in the long utterance was varied across subjects: For half the subjects the short utterance was located at the beginning of the long utterance, with the test items thus at the end of the long utterance, and for the other half the short utterance was located at the end of the long utterance, with the test items thus at the beginning. Consequently, the short utterance for some subjects was the test item for others, and vice-versa. In addition to the four item sets, two practice sets were included, in which the short utterances were monosyllabic and the long utterances were trisyllabic. The purpose of the practice sets was to generate practice test items, allowing subjects a few trials to get used to the task before presenting the experimental test items.

All utterances and test items were recorded by a female native speaker of English during a single session. She was instructed to pronounce each utterance slowly, at a regular pace and without pausing between syllables, placing the main stress on the first syllable of each utterance. This stress pattern was intended to minimize the possible influence of stress on segmentation. Furthermore, the design prevented an overall effect due to stress, since we predicted that different subject groups would segment identical utterance tokens in different ways.[4] Four different tokens of each long and short utterance were selected for presentation in the familiarization phase, adding some acoustic variety. The use of natural speech with several tokens of the same sequences of syllables was intended to make the recognition of a familiar unit more realistic: Recognizing, say, abc in the utterance abcde would imply an abstract mapping, since abc in isolation and abc in a longer utterance are acoustically different. In addition, stimulus variability appears to enhance spoken-word recognition (Nygaard, Sommers, & Pisoni, 1995).

Design

For each item set, four different familiarization conditions were created, crossing location (whether the short utterance was at the beginning or at the end of the long utterance) with length (whether this short utterance was two- or three-syllables long). These four conditions were presented to four different groups of subjects. Each subject received four item sets, one in each location-by-length condition, yielding a latin square design. For the test phase, each of the four groups was divided in half and each subgroup was tested on only one of the two possible test items for each item set (e.g., either cde or de). The order in which the test items from each set was presented was maintained constant across groups. This design (illustrated in Table 2) allowed us to present the same test item to two groups for which the predictions were different.

Insert Table 2 about here.

If differences in performance between the word and nonword conditions were found, they would result from the only thing that differed between the conditions, namely, which short utterance was presented during familiarization.

Procedure

Before familiarization, subjects were instructed that they would hear utterances consisting of one or more words from a "made-up" language. Speech materials was presented binaurally over headphones. The familiarization phase consisted of ten blocks. During each block, subjects heard all four short utterances and all four long utterances for their group, as well as four utterances from the practice item sets. In the first two blocks, each utterance was presented twice in a row and subjects were asked to repeat it aloud after each presentation. In the remaining eight blocks, each utterance was presented once and subjects were asked to listen and try to remember. Thus, each of the four short and four long utterances was presented 12 times, 4 with repetitions aloud and 8 without. The order of utterances within blocks was determined by generating a large number of random orders and discarding those in which two utterances from the same item set (e.g., ab and abcde) were presented adjacent to one another. Familiarization lasted about 12 minutes.

After familiarization, subjects heard “bits of speech” and had to decide for each whether it was a word of the made-up language.[5] To make the instructions as explicit as possible, an example was given, using new syllables: “For instance, if voguki is a word, chivoguki is not, nor is guki.” Subjects were asked to decide as quickly as possible, and to press one of two buttons labeled yes and no even if they were not quite sure of their response. The yes button was always controlled by the subject’s dominant hand. Each subject was presented with 6 “bits of speech”, two practice items followed by four test items. The response type and response latency for each item were collected by a computer, with reaction times measured from timing pulses aligned with item onset and inaudible to the subjects.

Results

Only response types will be reported here, since the response latencies did not show any effect. Test items in the word condition were judged to be a word (yes response) 62% of the time (59 out of 96) while in the nonword condition, they were judged to be a word 45% of the time (43 out of 96). Given the discrete nature of the data, a Chi-square test was required to test the difference between the two conditions directly. However, the Chi-square test requires that all data points be statistically independent, so it is not applicable to a table where some cells contain multiple responses from the same subject. We therefore computed the number of subjects who answered yes zero times, one time, and two times to items tested in the word condition. The maximum was two because each subject was tested on only two items in the word condition (and two in the nonword condition). This resulted in an exhaustive classification of subjects into three categories. We made the same computations for items presented in the nonword condition, yielding a second exhaustive classification of subjects. For each classification, we excluded subjects who failed to respond to one or more test items (one subject in the nonword condition and three subjects in the word condition). Table 3 presents the classifications of subjects by number of yes responses to test items in the word and nonword conditions.

Insert Table 3 about here.

A Chi-square test revealed the two classifications to be reliably different ($\chi^2_{(2)}$= 12.68, p<.002). This difference is reflected by the fact that more subjects gave two yes responses in the word condition than in the nonword condition (16 vs. 12 subjects), and more subjects gave zero yes responses in the nonword condition than in the word condition (16 vs. 4 subjects).[6]

Additional analyses were conducted, distinguishing the test items as a function of their previous location in the long utterance: at the left edge (e.g., *ab/abc* in *abcde*) or at the right edge (e.g., *cde/de* in *abcde*). Each subject was tested on two test items from each location, of which one was in the word condition and the other in the nonword condition. Table 4 presents the classifications of subjects by number of *yes* responses (0 or 1) to test items in the word and nonword conditions, broken down by location of the test item.

Insert Table 4 about here.

For each classification, we excluded subjects who failed to respond to the test item. Although the subject classifications for items in the nonword and word conditions differed significantly both for items on the left ($\chi^2_{(1)}$= 3.69, p=.055) and for items on the right ($\chi^2_{(1)}$= 10.53, p=.001), the effect of word vs. nonword seemed greater for items on the right. In order to test a possible interaction between the effect of the condition and the location of the test items, we computed for each subject the number of correct responses, that is, the number of *yes* responses to items in the word condition and *no* responses to items in the nonword condition, and computed subject classifications as a function of the number of correct responses (0, 1, or 2 correct responses) separately for the test items on the left and the test items on the right (see Table 5). As before, we excluded subjects who failed to respond to one or more items. The two distributions were not reliably different ($\chi^2_{(2)}$= 4.2, p=.12).[7]

Insert Table 5 about here.

We also classified subjects according to the number of yes responses (0 or 1) in nonword and word conditions broken down by number of syllables in the test item (Table 6). For each category, we excluded subjects who failed to respond to the test item. The subject classifications in nonword and word conditions differed significantly when the test items were two-syllables long ($\chi^2_{(1)}$= 15.96, p<.0001), but did not differ significantly when the test items were three-syllables long ($\chi^2_{(1)}$= 1.44, p=.23). In order to test a possible interaction between the effect of the condition and the number of syllables of the test items, we computed for each subject the number of correct responses, that is, the number of yes responses to items in the word condition and no responses to items in the nonword condition, and computed subject classifications as a function of the number of correct responses (0, 1, or 2 correct responses) separately for the 2-syllable test items and the 3-syllable test items (Table 7). As before, we excluded subjects who failed to respond to one or more items. The two distributions were not reliably different ($\chi^2_{(2)}$= 0.15).

Insert Tables 6 and 7 about here.

Discussion

Experiment 1 aimed to test whether a sequence of syllables which was presented in isolation (the short utterance) could be extracted from the edge of a longer utterance, yielding the inference of a new unit, the remainder of the utterance. We exposed different groups of subjects to the same long utterances but with short utterances of different length, yielding different predicted segmentation points and different predicted inferred units. Then we presented subjects with test items, and asked them to decide, for each test item, whether it was a word from the artificial language they had just heard. Across subjects, the same test items were either in the word or nonword condition, depending on the length of the short utterance subjects had been exposed to. The results showed a clear effect of the condition on the subjects' responses: Subjects classified a test item as being a word more frequently when it corresponded to the predicted inferred unit, showing a higher familiarity with this item than with a shorter or longer fragment of the long utterance. This suggests that the short utterance was extracted out of the long utterance, and that a new unit was isolated and stored in memory. Extraction of a familiar unit seems to have occurred both when located at the beginning and at the end of the long utterance.

However, this task appeared unsatisfactory in two ways. First, the task involved a metalinguistic judgment which was quite opaque for the subjects: Being a word for language users is much more than showing some degree of familiarity. Second, this task does not prevent a conscious strategy from playing a role. Despite the fact that they were not instructed about the task until after familiarization, some subjects could have noticed the short utterances occurring in the long utterances and later guessed that this pattern was related to what we meant when we asked them whether an item was a "word". Such a conscious guess at the meaning of the task is presumably not at work in early language acquisition. Experiment 2 thus replicated Experiment 1 using the same materials but a task that was not subject to conscious word-discovery strategies.

Experiment 2

The aim of Experiment 2 was to investigate whether subjects automatically segment long utterances by extracting a familiar unit at the edge and store the remainder of the utterance in memory, without requiring subjects to make metalinguistic judgments. For this purpose, we used a recognition-memory task. After a short familiarization like that of Experiment 1, subjects were presented with sequences of syllables and asked to decide, for each sequence, whether they had heard it during the familiarization phase. The sequences consisted of test items and distractors.

The test items were all fragments of the long utterances presented during familiarization, as in Experiment 1, so the correct answer was always yes. The distractors included sequences of syllables that had not been heard during familiarization, so the correct answer was no. As in Experiment 1, we describe test items as being in the word condition when they exactly match the remainder of the long utterance after the short utterance presented to the subject has been extracted from it; otherwise we describe them as being in the nonword condition. We expected subjects to show a high overall-accuracy level at distinguishing sequences of syllables they had heard during familiarization from those they had not. However, we had specific predictions about the pattern of errors they would make on the test items and about their speed at responding to them, that are independent of the overall accuracy. We predicted that subjects would be more accurate and perhaps faster at remembering that they had heard test items in the word condition as compared to those in the nonword condition, even though they had heard both the same number of times and in the exact same context within the long utterance. This prediction depends on the hypothesis that the representation of a sequence that has been stored as a unit is more accessible than the memory traces of sequences that have not been stored as a unit. This paradigm is less subject to conscious word-discovery strategies than the paradigm in Experiment 1, since the task does not refer to words at all.

Method

Subjects

Twenty-six students from Johns Hopkins University volunteered and were paid five dollars for their participation. They were all native speakers of English, and were tested individually in a sound-attenuated booth.

Materials

The materials were quite similar to the materials used in Experiment 1. The same four item sets were used. However, to simplify the design, the location of the short utterance in the long utterance (either at the left or the right edge) was varied across rather than within item sets: For the item sets *abcde* and *klmno*, the short utterance was located at the left edge, and for the item sets *fghij* and *pqrst*, at the right edge. For each item set, there were two short utterances of different lengths that were presented to different subject groups, and two test items of different length that were presented to all groups. In addition to the 8 test items (2 for each item set) for which the correct response was *yes*, 12 distractors were constructed for which the correct answer thus was *no*. They were of three types: Four were composed of two syllables drawn from different item sets (e.g., *ai*); four were composed of a pair of adjacent syllables and an additional syllable from the same set that never followed or preceded the pair during familiarization (e.g., *fgi*, *ade*); and four were composed of three syllables from the same set that never occurred successively (e.g., *jgi*, *adc*). To balance the number of correct *yes* and *no* responses in the test phase, four distractors were added that had been heard during familiarization. Two of these were bisyllabic and two trisyllabic. They were fragments of the long utterance that either contained or were contained in the short utterance — for example, *ab* for the group familiarized with *abc* and *abc* for the group familiarized with *ab.* The exact same materials as in Experiment 1 were used for the familiarization phase. To reduce the memory burden, only one of the two practice item-sets was used (/muzibo/, /bo/). All the materials presented in the test phase were recorded by the same female speaker as in Experiment 1, during a single session.

Design

The length of the short utterance for each item set was varied across two groups of subjects, yielding different segmentations and different predictions for the test items. As an example, the item set abcde is shown below, with an underscore indicating the predicted segmentation and boldface indicating the predicted inferred units in the test items:

	Familiarization	Test
Group 1	ab, ab_cde	**cde**, de
Group 2	abc, abc_de	cde, **de**

For the item set abcde for example, Group 1 heard the short utterance ab and the long utterance abcde, yielding the predicted segmentation ab_cde; conversely, Group 2 heard the short utterance abc and the long utterance abcde, yielding the predicted segmentation abc_de. These two groups yielded different predictions on the same test items. For instance, the test item cde corresponded to the inferred unit and was thus in the word condition for Group 1, while it straddled the segmentation point and was thus in the nonword condition for Group 2. This design allowed us to compare recognition performance for the same items in both conditions across groups. We predicted that performance would be better for test items in the word condition than in the nonword condition. Subjects were assigned to Group 1 for half the item sets, and to Group 2 for the other half.

Procedure

The familiarization phase was very similar to that of Experiment 1. In the test phase, subjects were asked to decide, for each sequence of syllables presented, whether they remembered it from the utterances they had just heard. An example was given, using sequences of syllables that did not occur in the materials. Subjects gave their response by pressing one of two buttons labeled yes and no, the yes button being always controlled by the subject's dominant hand. The instructions emphasized the speed of the response. Twenty-four items (8 test items and 16 distractors) were presented to each subject in a randomized order, constant across subjects.

Three practice test items were added at the beginning of the test phase, composed of syllables from the practice item set. Response type (yes or no) and reaction time for each item were collected by a computer, with reaction times measured from timing pulses aligned with item onset and inaudible to the subjects. Reaction time was collected during the period of 3500 msec after the onset of the test item. If the subject did not press one of the two buttons within this temporal window, the trial was considered to be missed.

Results

The overall accuracy for each subject (the proportion of correct responses to the test items and the distractors) was computed. The accuracies and reaction times for the test items only were then analyzed by an ANOVA. Throughout this study, the differences tested by items (F_2) were often found to be non significant. This is probably a consequence of the limited number of items used (eight). We did not add more test items because that would have meant presenting more utterances during familiarization, substantially increasing the memory burden and making the task more difficult. Throughout the analyses, our interpretation relies on the significance by subjects (F_1), although F_2 values are also reported.

The mean accuracy, that is, the proportion of correct responses to the test items and distractors together, was 73%. This indicates that subjects were able to accurately categorize a sequence of syllables as having been heard before or not 73% of the time. The performance on the distractors only was 72%; the distractors consisting of sequences of syllables that had not been heard before were responded to correctly 70% of the time, and the distractors consisting of sequences of syllables that had been heard before were responded to correctly 74% of the time.

Accuracy analysis

This analysis was restricted to the test items, all of which had been heard during familiarization and for which the correct response was *yes*. Percentages of accuracy for each subject and each condition were computed (*no* responses and missing data were coded as incorrect responses), and a 2×2×2 (condition by location by number of syllables) ANOVA by subjects (F_1) and by items (F_2) was conducted. Table 8 presents the accuracy percentages for each condition (nonword and word) as a function of the number of syllables of the test items.

Insert Table 8 about here.

The word/nonword condition was found to have a major impact on accuracy: The mean accuracy at responding to the test items was 75% in the nonword condition and 87% in the word condition ($F_1(1,25)=8.41$, $p<.01$; $F_2(1,4)=4.8$, $p=.09$). No global effect of the number of syllables or the location of the test item in the long utterance was found. None of the 2-way interactions among the nonword/word condition, location of test item, and number of syllables was significant.

Reaction-time analysis

RTs for test items that were accurately recognized (*yes* responses) were analyzed via a 2×2×2 (condition by location by number of syllables) ANOVA by subjects (F_1) and by items (F_2). Table 8 presents the mean reaction times for nonword and word conditions, as a function of the number of syllables of the test items. Mean RT was longer in the nonword condition (1578 ms) than in the word condition (1488 ms), but this effect was not significant ($F_1(1,25)=2.55$, $p=.12$; $F_2(1,4)=1.37$, $p=.31$). The 3-syllable test items showed longer mean RTs (1610 ms) than the 2-

syllable items (1463 ms), but the difference was only marginally significant ($\underline{F}_1$(1,25)=3.88, $\underline{p}$=.06; $\underline{F}_2$(1,4)=1.68, $\underline{p}$=.34). This difference was probably due to the duration difference between the 3-syllable items (1142 ms) and the 2-syllable items (794 ms), RTs being measured from the item onset. No global effect of the location of the test item was found. No interaction was found between the condition and the location of the test item or its number of syllables. No interaction between the number of syllables and the location of the test items was found either.

Discussion

Experiment 2, like Experiment 1, aimed to test whether a sequence of syllables that was presented in isolation (the short utterance) could be extracted from the edge of a longer utterance, yielding the inference of a new unit, the remainder of the utterance. By contrast with Experiment 1, we used an indirect measure to account for the discovery and storage of new units: subjects' accuracy and speed at remembering a sequence of syllables they had heard during familiarization. After a short familiarization phase, we presented subjects with test items, and asked them to decide whether they remembered having heard the sequences of syllables before. Subjects were more accurate at remembering sequences of syllables when these sequences corresponded to the remainder of the utterance after extraction of a familiar unit than when the sequences corresponded to longer or shorter utterance fragments.

Experiment 2 supports INCDROP' s predictions; namely, that a familiar unit is extracted from the edge of an utterance and the remainder of the utterance is inferred as a new unit. The inference and storage of new units after segmentation of the familiar unit was demonstrated indirectly, by testing subjects' performance at accessing the memory traces of sequences of syllables. If the sequence did not correspond to an inferred unit, its retrieval was more prone to failure than the retrieval of a sequence that corresponded to an inferred unit. This suggests that new units can be represented in memory without having been presented in isolation, through the extraction of familiar units. Such a word-discovery mechanism therefore seems to exist and to be automatically applied by listeners engaged in a simple memorization task.

The next question, addressed in Experiment 3, was whether the extraction of a familiar unit embedded in the middle of an utterance could be observed under similar conditions.

Experiment 3

Experiments 1 and 2 demonstrated that people tend to extract a familiar unit from the edge of an utterance, treating the remainder of the utterance as a novel unit. INCDROP favors this segmentation because it reduces the total length of all novel units in the segmentation, as compared to the alternative in which the entire utterance is treated as a single novel unit. Further, when the familiar unit is at the edge, the number of novel units is one regardless of whether the familiar unit is segmented out or the whole utterance is treated as a long novel unit. However, when a familiar unit is embedded in the middle of an utterance containing no other familiar units, extracting the familiar unit implies two novel units, one to the left of the familiar unit and one to the right. As a result, INCDROP predicts that extracting a familiar unit and inferring the new units should be more difficult when the familiar unit is embedded in the middle of a long utterance, although it should still be possible if the familiar unit is long enough and frequent enough.

There is evidence in the language-acquisition literature that the processing of a word in utterance-medial position is harder than in utterance-final position. Young children's comprehension of words seems to be better when the words are in utterance-final position than when they are in utterance-medial position (Fernald, McRoberts, & Herrera, in press; Swingley & Pinto, 1997). Furthermore, adult language learners have been shown to be more efficient at recognizing a word when it occurs at the end of an utterance than when it is embedded in the utterance (Golinkoff & Alioto, 1995). Interestingly, English adult speakers tend to place focused words conveying new information at the end of utterance in both adult- and child-directed speech, which might be an efficient strategy to optimize the processing of the most informative part of the utterance (Aslin, 1993; Fernald & Mazzie, 1991).

Experiment 3 was designed to test whether a familiar unit embedded in the middle of a long utterance would be extracted and new units inferred, using the same paradigm as in Experiment 2. We manipulated the short utterances that subjects heard during familiarization in order to see how that would affect the new units they inferred from the long utterances. A sample item set is shown below, with the predicted inferred unit for each group in bold:

	Familiarization	Test
Group 1	bc, a_bc_de	cde, **de**
Group 2	ab, ab_cde	**cde**, de

Subjects in Group 1 were presented with a short utterance, bc, which later occurred embedded in the middle of the long utterance abcde. We hypothesized that the short utterance would be recognized in the long utterance, extracted, and that two new units would be inferred, a and de. In order to assess the extraction of the new unit de, we needed to compare subjects' performance on the test item de with other subjects' performance on the same test item when it did not correspond to an inferred unit. Subjects in Group 2 were thus familiarized with the short utterance ab and the long utterance abcde; based on the results from Experiments 1 and 2, the long utterance was predicted to be segmented and the new unit cde to be inferred. For the subjects in Group 2, the test item de did not correspond to the inferred unit, and their ability to remember having heard this item was predicted to be lower that for the subjects in Group 1. The same reasoning was held for the test item cde. This design allowed us to compare performance on the same test items across groups as a function of condition.

Method

Subjects

Twenty-four students from University of Maryland, College Park, volunteered and got course credit for their participation. They were all native speakers of English, and were tested individually in a quiet room.

Materials

Four item sets were used. The syllabic composition of the item sets was different from Experiments 1 and 2. The long utterances were /digufezomu/ (abcde), /potekobudo/ (fghij), /kubegifoze/ (klmno), and /rubɔ:nevoli/ (pqrst). Note that the letter code used is similar to the previous experiments. Each item set was composed of one long utterance (e.g., abcde), two bisyllabic short utterances, one located at the edge of the long utterance (e.g., ab), the other embedded in the long utterance (e.g., bc), and two test items consisting of syllables from the part of the long utterance that did not contain the short utterances (e.g., cde and de). The location of the short utterance, either at or towards the beginning of the long utterance or at or towards the end, was varied across item sets. To reduce the memory burden, no practice item set was used, since subjects missed very few test items in Experiment 2. In addition to the eight test items (two for each of four item sets), eight distractors were constructed, composed of syllables in new sequences; four were bisyllabic, four trisyllabic. There were thus eight test items consisting of sequences that occurred during familiarization and for which the correct response was yes, and eight distractors consisting of sequences that did not occur during familiarization and for which the correct response was no. The number of 2-syllable and 3-syllable items was equal. The materials were recorded by the same female speaker as in the previous experiment, and as before, she was instructed to produce each syllable at a regular pace, with the main stress on the first syllable of the utterance.

Design

For each item set, the type of the short utterance (at the edge or embedded) was varied across two groups of subjects. For the item set abcde for example, Group 1 heard the short utterance bc and the long utterance abcde, predicting the segmentation a_bc_de; conversely, Group 2 heard the short utterance ab and the long utterance abcde, predicting the segmentation ab_cde. In the test phase, the test item de was in the word condition for Group 1 but in the nonword condition for Group 2, while the test item cde was in the nonword condition for Group 1 and in the word condition for Group 2. We thus compared performance on the same test items in different conditions, across groups. Subjects were assigned to Group 1 for half the item sets, and to Group 2 for the other half.

Procedure

Before the familiarization phase, the same instructions as in Experiment 2 were given to subjects. The familiarization phase consisted of 30 blocks during which each of the four short and four long utterances were presented. This increase in repetitions compared to Experiment 2 was intended to increase the probability that the short utterance would be remembered, recognized, and extracted. As in Experiment 2, the first two blocks were repetition blocks and the remainder were simple listening blocks. The order of utterances within blocks was determined by generating a large number of random orders and discarding those in which two utterances from the same item set (e.g., ab or bc and abcde) were presented adjacent to one another. Familiarization was about 20 minutes long. Subjects were given a short break after 12 minutes of listening, to provide them with a rest, and also to re-stimulate their interest half way through the listening. The procedures for the test phase and the data collection were identical to those of Experiment 2.

Results

The overall accuracy on test items and distractors was 74% on average. For the distractors only, the accuracy was 73%.

Accuracy analysis

The accuracy percentages on the test items only were submitted to a 2×2×2 (condition by location by number of syllables) ANOVA by subjects ($\underline{F}_1$) and by items ($\underline{F}_2$). Table 9 presents the mean percentages.

Insert Table 9 about here.

The accuracies at responding to items in the word and nonword conditions did not differ significantly (76% and 74% respectively). No effect of the location of the test item in the long utterance (left or right) was found. Accuracy differed between 2- and 3-syllable items (57% vs. 93%, $\underline{F}_1(1,23)=39.8$, $\underline{p}<.0001$; $\underline{F}_2\ (1,4)=10.91$, $\underline{p}<.05$). No interaction between the condition and the location of the item was found, nor between the condition and the number of syllables or between the number of syllables and the location of the test items.

Reaction-time analysis

Table 9 also presents the mean reaction times for the yes responses in the nonword and word conditions. These results were analyzed in a 2×2×2 (condition by location by number of syllables) ANOVA by subjects ($\underline{F}_1$) and by items ($\underline{F}_2$). No significant difference was found between test items in the nonword condition and in the word condition. The 2-syllable items were responded to significantly faster than the 3-syllable items (1623 ms vs. 1732 ms, $\underline{F}_1$(1, 23)=9.4, p<.01; $\underline{F}_2$(1,4)=1.43, p=.30), as a probable consequence of the difference in duration of the test items (762 ms and 1047 ms on average, respectively). No global effect of the location of the test items in the utterance was found. No interactions were found between the condition and the number of syllables or the location of the test item in the long utterance.

Discussion

The accuracy and response-latency analyses failed to show any difference between test items in the word and nonword conditions. Thus, we failed to find any evidence that subjects inferred new units by extracting a familiar unit from the midst of an utterance. Although we cannot take this as a confirmation of the null hypothesis, it raises the possibility that the subjects in fact did not infer new units in this way.

The one reliable effect in this experiment was that subjects remembered having heard 3-syllable test items more often than 2-syllable test items (93% versus 57%). There are two compatible explanations for this. First, it is possible that 3-syllable items in general convey more sequential information than 2-syllable items, and therefore were less prone to errors. For example, a subject who recognizes one pair of adjacent syllables in a 3-syllable item may be able to use that pair to retrieve his or her memory of the entire 3-syllable item by association; for two syllable items there is no second chance. Likewise, a strongly unfamiliar pair may be enough to cause subjects to reject a 3-syllable distractor they have not heard before. This interpretation is supported by the accuracy advantage for 3-syllable distractors, compared to the 2-syllable distractors , although this difference was only marginally significant (78% vs. 68%, $\underline{F}_1$(1,23)=3.44, p=.08; $\underline{F}_2$<1). A second explanation for subjects' greater tendency to recognize 3-

syllable test items is that these items contain a syllable that occurred more frequently during the familiarization phase than the other syllables of the test items. For example, when bc and abcde were presented during familiarization, the syllables b and c occurred twice as often as the syllables a, d, or e. Test items that contained a high-frequency syllable (c in the test item cde or m in the test item klm) were 3-syllables long; they might have been responded to with a higher accuracy because the frequent syllable served as an anchor, activating the memory traces of the sequence where it appeared more efficiently than the less frequent syllables.

The lack of a reliable difference between the word and nonword conditions in Experiment 3 could be due to acoustic factors (see General Discussion). Future experiments using more sensitive methods may reveal that subjects can infer novel words by extracting a familiar word from the middle of an utterance. Nonetheless, the results of Experiment 3 suggested the possibility that subjects might have more difficulty inferring novel word-like units by extracting familiar units from the middle of an utterance than by extracting familiar units from the edge. Experiment 4 investigated that possibility directly.

Experiment 4

Experiment 4 was designed to show a positive effect if subjects had more difficulty inferring new words by extracting bc from abcde than by extracting ab from abcde. In this experiment, one group of subjects (Group 1) was presented with two short utterances, ab and bc, and a long utterance, abcde. Since these two short utterances share a syllable (b), they could not both be extracted within the same segmentation and hence they competed for the segmentation of the long utterance: The extraction of ab would yield the segmentation ab_cde, while the extraction of bc would yield the segmentation a_bc_de. INCDROP predicts that ab, rather than bc, should be extracted, because the extraction of ab minimizes the number of novel words. If the extraction of ab is favored against the extraction of bc, the new unit cde should be inferred and stored in memory. In order to assess this segmentation, we needed to compare the performance on the test item cde for this group of subjects with the performance of another group, for which that test item would not correspond to an inferred unit. This second group of subjects (Group 2) was thus presented with the short utterance abc which should trigger the segmentation abc_de. For this group, the test item cde is not an inferred unit, and thus should be remembered less well than for Group 1; conversely, the test item de should be remembered better for Group 2 than for Group 1, for which it is not an inferred unit. The design and predictions for the item set abcde are shown below, with predicted inferred units in bold, on the assumption that subjects in Group 1 would extract the familiar unit from the edge rather than the middle of the utterance:

	Familiarization	Test	
		Inference	Control
Group 1	ab, bc, ab_cde	**cde**, de	abc, **ab**
Group 2	abc, abc_de	cde, **de**	**abc**, ab

As in Experiments 2 and 3, we compared subjects' performance at remembering predicted inferred units to their performance at remembering items that were longer or shorter than the predicted inferred unit (inference test items). As a control, we also compared their performance at remembering the short utterances they heard in isolation (e.g., ab for Group 1, abc for Group 2) to their performance at remembering items that were one syllable longer or shorter (control test items). This control was meant to ensure that a sequence of syllables is remembered better when it corresponds to a unit in memory than when it does not, using a simpler test case where the sequence of syllables was bounded by silence during familiarization.

Method

Subjects

Twenty-three students from University of Maryland at College Park volunteered and got course credit for their participation. They were all native speakers of English, and were tested individually in a quiet room.

Materials

Four item sets were used, the same ones as in Experiment 3 (but different from those used in Experiment 2). Each item set was composed of a long utterance (e.g., abcde), two bisyllabic short utterances (e.g., ab, bc), one trisyllabic short utterance (e.g., abc), and two tests items (e.g., cde, de). Distractors consisted of bisyllabic and trisyllabic sequences of syllables that did not occur in sequence during the familiarization phase. We used the tokens from Experiment 3, supplemented with new trisyllabic short utterances and additional distractors. The same female speaker recorded the new materials, in a single session. In addition, one more token of the short utterances was selected from the previous recording for the control test items (e.g., ab), so none of the test items in the test phase were acoustically identical to what had been heard during the familiarization. There were 16 test items consisting of syllable sequences heard during familiarization and 16 distractors consisting of syllable sequences that were not heard during familiarization. Of the 16 test items, 8 were intended to test whether recognition was facilitated for inferred units as compared to items that were shorter or longer by one syllable (inference test items); the other 8 were controls to check that recognition was facilitated for the short utterances, as compared to items that were shorter or longer by one syllable (control test items). Out of these 32 items, the number of trisyllabic and bisyllabic items was counterbalanced, as well as the number of times each syllable was presented across the 32 items.

Design

For each item set, one group of subjects heard the two overlapping short utterances and another group heard the single trisyllabic short utterance. Both groups heard the same long utterance. For the item set abcde for example, Group 1 heard the short utterances bc and ab and the long utterance abcde, for which we predicted the segmentation ab_cde; Group 2 heard the short utterance abc and the long utterance abcde, for which we predicted the segmentation abc_de. In the test phase, the test item cde was in the word condition for Group 1 but in the nonword condition for Group 2; conversely, the test item de was in the word condition for Group 2 but in the nonword condition for Group 1. We compared the subjects' performance on the same items, across groups, and we predicted better performance for test items when they were in the word condition than in the nonword condition. Such a difference would indicate that the two groups had inferred different new units, and therefore that the extraction of ab was favored against the extraction of bc, as predicted. Subjects were assigned to Group 1 for half the item sets, and to Group 2 for the other half.

Procedure

The procedure was identical to that of Experiments 2 and 3. During the familiarization phase, each (long and short) utterance of each item set was presented in two repetition blocks followed by 28 simple listening blocks. The duration of the familiarization was about 24 minutes, with a break after 14 minutes.

Results

The overall accuracy on distractors and test items together was 77% on average. Subjects were thus able to accurately remember whether a sequence of syllables had been heard before or not 77% of the time. The accuracy on distractors only was also 77%.

Inference test items

Table 10 presents the accuracies and reaction times for inference test items in nonword and word conditions as a function of the number of syllables in the item. The accuracy and reaction times were submitted to separate 2×2×2 (condition by location by number of syllables) ANOVAs by subjects (F_1) and by items (F_2).

Insert Table 10 about here.

Accuracy analysis. Accuracy was higher for test items in the word condition than in the nonword condition (73% vs. 60%, $F_1(1,22)=5.35$, $p<.05$; $F_2(1,4)=2.74$, $p=.17$). Accuracy for 3-syllable items was higher than for 2-syllable items (83% vs. 50%, $F_1(1,22)=26.19$, $p<.0005$; $F_2(1,4)=3.45$, $p =.14$). No effect of the test-item location in the utterance was found. No significant interaction between the condition and the number of syllables or the location, or between the number of syllables and the location of the test item was found.

Reaction-time analysis. Items in the word condition were responded to faster than in the nonword condition (1613 ms vs. 1791 ms), but the difference was only marginally significant ($F_1(1,21)=3.81$, $p=.06$, the mean squared error being computed on 22 subjects, because of missing data; $F_2<1$). Three-syllable items showed longer mean RT than 2-syllable items (1777 ms vs. 1554 ms, $F_1(1,21)=8.42$, $p<.01$; $F_2(1,4)=7.51$, $p=.05$), almost certainly due to their duration difference. No global effect of the location of the test items in the long utterance was found, and no interaction between condition and number of syllables or location, or between the number of syllables and the location either.

Control test items

Table 10 also presents accuracy percentages and RTs for the control test items in nonword and word conditions as a function of the number of syllables in the item. Accuracy percentages and RTs were submitted to separate 2×2×2 (condition by location by number of syllables) ANOVAs by subjects (F_1) and by items (F_2).

Accuracy analysis. Control test items in the word condition were responded to more accurately than in the nonword condition (98% vs. 88%, $F_1(1,22)=5.75$, $p<.05$; $F_2(1,4) < 1$). No effect of the number of syllables or the location in the long utterance was found. No significant interaction between the condition and the number of syllables or the item location, or between the number of syllables and location of the test item was found.

Reaction-time analysis. Control test items were responded to faster in the word condition than in the nonword condition (1397 ms and 1548 ms respectively, $F_1(1,22)=9.70$, $p<.005$; $F_2(1,4)=2.21$, $p=.21$). Moreover, 2-syllable items were responded to faster than 3-syllable items (1334 ms and 1600 ms respectively, $F_1(1,22)=42.53$, $p<.00001$; $F_2(1,4)=5.22$, $p=.08$), which again can easily be explained by the duration difference between these items (820 ms and 1120 ms, respectively). No effect of the location of the item in the long utterance was found. No interaction between the condition and the location of the test item was found, nor between the number of syllables and location.

Comparison between inference and control test items

The accuracies and reaction times to both the control test items and the inference test items

were analyzed together in a new 2x2x2x2 ANOVA that included test-item type (control vs. inference) as a factor.

were analyzed together in an new, 2×2×2×2 ANOVA that included test-item type (control versus inference) as a factor. Accuracy was higher for control than inference test items (on average 93% vs. 66%, $\underline{F}_1(1,22)=53.9$, $\underline{p}<.0005$; $\underline{F}_2(1,8)=8.99$, $\underline{p}<.05$). A word-nonword difference in accuracy was found both for the control and inference test items (9.8% and 13%, respectively), with no interaction between the condition and the test-item type (control or inference). As far as the response latencies were concerned, control test items were responded to faster than inference test items (1468 ms vs. 1693 ms, $\underline{F}_1(1,22)=16.9$, $\underline{p}<.001$; $\underline{F}_2(1,8)=7.28$, $\underline{p}<.05$). Both control and inference test items were responded to faster in word condition than in nonword conditions (by 151 ms and 178 ms, respectively), with no interaction between the condition and the test-item type (control or inference).

Discussion

Experiment 4 aimed to determine the outcome of a competition between a familiar unit occurring at the edge of the utterance and a familiar unit embedded in the middle of the utterance. INCDROP explicitly predicts that the segmentation involving the extraction of a familiar unit at the edge of the utterance should be favored against the segmentation involving the extraction of a familiar unit embedded in the utterance, since the former implies only one new unit while the latter implies two. To test this prediction, we compared the responses to test items in word and nonword conditions, as predicted by the model. Test items in the word condition were responded to more accurately, and tended to be responded to faster than items in the nonword condition, supporting the model's predictions. However, we do not have evidence that the two familiar units actually competed for segmentation. It is possible that the medially-embedded familiar unit was perceptually more difficult to recognize and extract than the familiar unit at the edge. We return to this point in the General Discussion. As in Experiment 3, accuracy for 3-syllable inference test items was higher than for 2-syllable inference test items. The presence of a high-frequency syllable and/or the more informative nature of 3-syllable sequences could explain this effect. The fact that subjects were able to recognize that they had not heard 3-syllable distractors better than 2-syllable distractors supports the latter hypothesis (86% vs. 65%, $\underline{F}_1(1,22)=24.38$, $\underline{p}<.0005$; $\underline{F}_2(1,14)=5.64$, $\underline{p}<.05$).

In addition to testing the inference test items, we tested the recognition of control test items, which consisted of either the exact short utterance heard in isolation or a sequence that was longer or shorter by one syllable. Accuracy was higher when the control test items corresponded to the exact short utterance. The response latencies showed a clear facilitation for responding to the control test items in the word condition compared to the nonword condition. This nonword-word effect for the control test items confirms that the accuracy and response latency at remembering a sequence of syllables are enhanced when the sequence consists of a unit in memory; both the short utterance bounded by silence and the new unit inferred after segmentation showed such a facilitation.

General Discussion

Summary of predictions and results

INCDROP is a model put forth to explain how humans at all stages of linguistic knowledge discover new word-like units from continuous speech (Brent, 1997; see also Brent 1996, Brent & Cartwright, 1996). It posits a single mechanism that discovers new units by recognizing familiar units in an utterance, extracting those units, and treating the remaining contiguous stretches of the utterance as novel units. When an utterance contains no familiar units the whole utterance is treated as a single novel unit, so there is no need to assume a special bootstrapping device that discovers the first units. Like all the proposed phonological strategies for segmentation and word discovery, this mechanism is not expected to be foolproof. Rather, the theory is that this mechanism generates a set of units that are candidates for becoming the sound patterns of words, if an appropriate meaning and/or syntactic function can be assigned to them.

This series of experiments suggest that new units can be isolated by recognizing familiar units and segmenting them out of the longer context they are embedded in. The evidence is that subjects showed higher familiarity and stronger memory traces for sequences of syllables that correspond to the new units predicted by this process, as compared to shorter or longer subsequences of the context. However, this does not imply that the new units have acquired the status of words for the subjects, nor that they would automatically acquire lexical status for child learners presented with analogous linguistic stimuli. Conversely, the sequences that violate unit boundaries are not necessarily excluded from acquiring lexical status. Indeed, subjects consistently found these latter sequences more familiar than sound patterns formed by syllables that did not occur successively, as shown by the high percentages of yes responses for test items in the nonword condition in all experiments presented here. Yet, subjects treated sequences of syllables in the word and nonword conditions differently. We interpret subjects’ greater sense of familiarity with the items in the word condition as evidence that these items are better candidates than items in the nonword condition for binding with meaning and/or syntactic function to form lexical entries.

At a more detailed level, INCDROP asserts that people segment utterances in such a way as to minimize the number of implied novel word-like units, minimize the length of the utterance attributed to novel units, and maximize the product of relative frequencies of the segmented units. Brent and Cartwright (1996) showed, using computer simulations, that segmenting according to these criteria was effective at extracting words from broad phonetic transcription of child-directed speech. The present study aimed to test whether listeners actually segment utterances and infer new units as predicted by the model. In the experiments reported here, adult subjects were exposed to short and long utterances from a miniature artificial language. Some of the short utterances appeared at the edge of a long utterance and others appeared in the middle of a long utterance. The model predicts that the short utterance will become a familiar unit. Further, when a familiar unit occurs at the edge of the long utterance, the model predicts that it will be extracted and that the remainder of the long utterance will be treated as a novel unit, unless the familiar unit is both very short and very rare. When the familiar unit occurs in the middle of the long utterance, however, segmenting it out implies two novel units, while treating the whole utterance as a novel unit implies only one. This cost in more novel words also implies a cost in lowering the product of relative frequencies. Thus, the model predicts that extracting a familiar unit from the middle of a longer utterance will be more difficult than extracting from the edge, although extracting from the middle should still be possible if the familiar unit is sufficiently long and sufficiently frequent.

Experiments 1 and 2 tested the inference of a new unit after extraction of a familiar unit located at the edge of the long utterance. When asked to judge whether a test item was a word of the made-up language (Experiment 1), subjects considered the predicted inferred units to be words more often than fragments of the long utterance that were longer or shorter than the predicted inferred units. When asked whether they had heard a sequence of syllables during the familiarization phase (Experiment 2), subjects were more accurate at recognizing the predicted inferred units than fragments of the long utterance that were longer or shorter than the predicted inferred units. Experiment 3 aimed to test whether a familiar unit can be extracted from the middle of the long utterance, causing two new units to be inferred. One group of subjects was exposed to a short utterance that appeared at the edge of the long utterance (e.g., ab and abcde) while another group was exposed to a short utterance that appeared in the middle of the long utterance (e.g., bc and abcde). If both groups segmented out the short utterance and treated the remaining contiguous stretches as new units, we would expect the first group to treat cde as a unit but not de, and vice-versa for the second group. We did not find a reliable difference between the two groups. This led us to design Experiment 4 in such a way that there would be an effect if subjects found it easier to extract a familiar unit from the edge of an utterance than from the middle. In this experiment, we presented both an edge-embedded familiar unit and a medially-embedded familiar unit to the same subjects. Both familiar units could not be extracted in the same segmentation because they overlapped by one syllable. The results suggested that the familiar unit at the edge, rather than in the middle, was extracted and the remainder of the utterance was inferred as a new unit.

These results with adult subjects suggest that the human mind is capable of the types of computations and behaviors predicted by INCDROP, but experiments with young children would be required to prove that the same conclusions apply to children in particular.

Possible explanations for the difficulty of extracting from utterance-medial position

The failure to find extraction of a familiar unit from the middle of an utterance may be due to perceptual factors, cognitive factors, or both. One perceptual explanation is that syllables at utterance boundaries are more acoustically salient. Words tend to be longer in utterance-final position, and pitch excursions larger in utterance-initial position (Klatt, 1974, 1975, 1976; Oller, 1973; Umeda, 1977). Further support for the influence of utterance position on salience comes from Fernald et al. (in press), who found that 15-month-old infants showed notably enhanced comprehension of utterance-medial words when their duration was increased. A second perceptual explanation depends on the degree of acoustic similarity between the short utterance when it occurs in isolation and the same syllable sequence when it occurs within the longer utterance. Consider once again the fact that words tend to be longer in utterance-final position, and pitch excursions are larger in utterance-initial position. When bc occurs in isolation, its two syllables are initial and final, respectively; when it occurs in abcde both of its syllables are medial. When the short utterance occurs at the edge of the long utterance, however, one of its syllables is in the same position as when the short utterance occurs in isolation. In addition, co-articulation may render medially-embedded utterances less acoustically similar to their isolated counterparts. When the short utterance is at the edge of the long utterance, only the medial end of the short utterance (e.g., b when ab is embedded in abcde) is subject to influence by neighboring segments; when the short utterance is in the middle of the long utterance, both ends are.

The relative difficulty of extracting a familiar unit from the middle of an utterance as compared to the edge could also be explained by cognitive factors. Descriptively, better recognition of familiar units at the edges of the utterance could be attributed to recency and primacy effects. However, INCDROP explains this effect in terms of the number and length of novel units — familiar units at the edges of utterances simply happen to imply fewer novel units. The INCDROP explanation therefore makes additional predictions, as described below.

The experiments presented here do not even distinguish between the perceptual and the cognitive account, much less between the various versions of each. However, one can imagine future work that might, in principle, show that the advantage of edge-embedded units over medially-embedded units is at least partially explained by INCDROP. A first step would be to relieve the potential perceptual difficulties by making the embedded and isolated occurrences of the short utterance acoustically identical. The chief disadvantage of this approach, and the reason it was not used here, is that the result inevitably sounds unnatural, and that the unnaturalness may well wipe out any improvement in recognizability due to acoustic similarity. Another step would be to design materials that could distinguish the INCDROP explanation, which depends on the number of implied novel units, from the perceptual and recency/primacy explanations, which depend directly on the edge of the long utterance. Informally, INCDROP favors extracting familiar units that occur adjacent to other familiar units as well as those that occur at the edge of the utterance. For example, suppose that ab, cd, and de are all familiar units. INCDROP predicts that the utterance abcdef will be segmented as ab cd ef, with the two familiar units ab and cd adjacent to one another, not as ab c de f, with the two familiar units ab and de separated by the novel unit c. The reason is that the segmentation where the familiar units are adjacent minimizes the number of novel units. In this example, the preference for fewer novel units does not coincide with a preference for extracting familiar units at the edges of the utterance. The chief difficulty of this approach is that both the cognitive and perceptual burdens in this design are substantially greater than those in the designs used so far. Specifically, subjects must learn three familiar units and recognize two of them in a larger utterance, including one in medial position. Finally, it should be noted that all of the explanations laid out above could contribute to the results we have observed; finding effects that can only be explained by INCDROP would not imply that the other factors cited above play no role.

Could these results be explained by transitional probabilities alone?

The results reported are consistent with predictions of the INCDROP model, but we must consider whether they could be explained equally well by a segmentation procedure based solely on transitional probabilities. In order to answer this question, we must assume a precisely formulated account of transitional-probability segmentation. For this purpose, we adopt the mathematical definition of transitional probability offered by Saffran, Newport, and Aslin (1996), and we assume that segmentation occurs precisely when the transitional probability between two adjacent syllables is lower than the transitional probability between the pair of syllables on either side. The analysis presented here does not apply to any other segmentation algorithm based on transitional probabilities or any other statistics on the co-occurrence of adjacent syllables that might be proposed.

Saffran, Newport, and Aslin (1996) define the transitional probability between two syllables *x* and *y* as the proportion of occurrences of *x* that are followed by *y*. Clearly, transitional probabilities are asymmetric, taking account of the proportion of *x*'s followed by *y* but not the proportion of *y*'s preceded by *x.* One consequence of this asymmetry is that transitional probabilities make the same prediction as INCDROP for Experiments 1 and 2 when the familiar unit is at the beginning of the long utterance, but not when it is at the end. For example, consider subjects exposed to *ab* and *abcde* with equal frequency. The short utterance *ab* is at the beginning of *abcde*, and the transitional probability is lowest between *b* and *c* — half of the occurrences of *b* were followed by a pause, half by the syllable *c*, while the other syllables were consistently followed by the same context. Now consider subjects exposed to *de* and *abcde* with equal frequency. The short utterance *de* is at the end of *abcde*, and the transitional probabilities in the long utterance are all equally high, since each syllable was always followed by the same context. Although *d* is *preceded* by both *c* and silence, it is only the following context that determines transitional probability. Thus, the transitional probabilities do not predict any segmentation when the short utterance is at the end of the long utterance. However, we found segmentation and inference with short utterances at either edge.

In Experiment 4 we found evidence that subjects exposed to *ab*, *bc*, and *abcde* with equal frequency infer that *cde* is a novel unit. This implies that they segmented the utterance between *b* and *c*, as predicted by INCDROP. The only local minimum of transitional probability, however, lies between *c* and *d.* To see this, note that all the *a*’s are followed by *b*, two-thirds of the *b*’s are followed by *c*, only half of the *c*’s are followed by *d*, all *d*’s are followed by *e*, and all *e*’s are followed by silence.

These observations seem to suggest that a segmentation strategy based only on transitional probabilities does not accurately predict patterns of segmentation when the input consists of short utterances. Nonetheless, transitional probabilities might work together with lexically driven segmentation. For example, they might serve as a useful measure of whether a hypothesized new word “sounds like” a word of the language being learned — that is, whether it has probabilistic phonotactics appropriate for that language. Further, INCDROP’s prediction that utterances will be stored as a single unit when they contain no familiar units breaks down when all input utterances are very long. INCDROP only models the relative cognitive burden of different ways in which an utterance could be exhaustively divided into familiar words and novel words to be memorized; it does not predict what will happen when the burden becomes so great that it is not possible to process the utterance exhaustively and some parts of it must be abandoned, neither recognized as familiar units nor learned as novel units. In such cases, local statistics such as transitional probabilities may play a dominant role in segmentation.

Language and memory

The work reported here began with a theory about how children acquire language, and its ultimate aim is still to improve our understanding of language acquisition. However, INCDROP can be thought in more general, cognitive terms, as a model of how people divide up speech and store it in memory. Indeed, what we have shown in these experiments is more directly interpretable in terms of general cognitive strategies than in terms of native-language acquisition. Nonetheless, if children have at their disposal the same types of automatic storage strategies as adults, that could help explain how they accomplish one of the many remarkable feats of language acquisition. One need not be an empiricist about syntax and phonology to believe that word discovery piggy-backs on general strategies for storing sequences in memory.

References

Aslin, R. N. (1993). Segmentation of fluent speech into words: learning models and the role of maternal input. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), Developmental neurocognition: Speech and face processing in the first year of life (pp. 305-315). Dordrecht: Kluwer Academic Publishers.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), Signal to syntax: Bootstrapping from speech to grammar in early acquisition (pp. 117-134). Mahwah, NJ: Lawrence Erlbaum Associates.

Bernstein Ratner, N. (1996). From “signal to syntax”: but what is the nature of the signal? In J. L. Morgan & K. Demuth (Eds.), Signal to syntax: Bootstrapping from speech to grammar in early acquisition (pp. 135-150). Mahwah, NJ: Lawrence Erlbaum Associates.

Bower, G. H., & Springston, F. (1970). Pauses as recoding points in letter series. Journal of Experimental Psychology, 83, 421-430.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: effects of item and pattern frequency, arbitrary subclasses, and correction. Journal of Memory and Language, 29, 591-610.

Brent, M. R. (1996). Advances in the computational study of language acquisition. Cognition, 61, 1-37.

Brent, M. R. (1997). Toward a unified model of lexical aquisition and lexical access. Journal of Psycholinguistic Research, 26, 363-375.

Brent, M. R. (in press). An efficient, probabilistically sound segmentation algorithm. To appear in Machine Learning Journal.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. Cognition, 61, 93-125.

Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: the contribution of phonological markers to learning. Journal of Memory and Language, 32, 76-95.

Brown, R., & Bellugi, U. (1964). Three processes in the child's acquisition of syntax. Harvard Educational Review, 34, 133-151.

Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 550-566.

Christiansen, M. H., Allen, J., and Seidenberg, M. S. (in press). Learning to segment speech using multiple cues: A connectionist model. Language and Cognitive Processes.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. T. M. Altmann (Ed.), Cognitive models of speech processing. Psycholinguistics and computational perspectives (pp. 105-121). Cambridge, MA: MIT Press.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. Lingua, 92, 81-104.

Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan & K. Demuth (Eds.), Signal to syntax: Bootstrapping from speech to grammar in early acquisition (pp. 87-99). Mahwah, NJ: Lawrence Erlbaum Associates.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. Computer Speech and Language, 2, 133-142.

Cutler, A., & Norris, D. (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., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. Journal of Memory and Language, 36, 202-225.

Echols, C. H., & Newport, E. L. (1992). The role of stress and position in determining first words. Language Acquisition, 2, 189-220.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. Journal of Acoustical Society of America, 97, 1893-1904.

Fernald, A., McRoberts, G., & Herrera, C. (in press). Effects of prosody and word position on lexical comprehension in infants. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. Developmental Psychology, 27, 209-221.

Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: initial representations for language learning. In F. J. Newmeyer (Ed.), Linguistics: the Cambridge survey. Volume III: Language: psychological and biological aspects (pp. 150-193).Cambridge: Cambridge University Press.

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: the state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), Language acquisition: the state of the art (pp. 3-48). New York: 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. Journal of Child language, 20, 229-252.

Hakuta, K. (1976). Becoming bilingual: A case study of a Japanese child learning English. Language Learning, 26, 321-351.

Hayes, J. R., & Clark, H. H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. R. Hayes (Ed.), Cognition and the development of language (pp. 221-234). New-York, John Wiley & Sons.

Hickey, T. (1993). Identifying formulas in first language acquisition. Journal of Child Language, 20, 27-41.

Houston, D., Jusczyk, P. W., & Newsome, M. (1995, November). Infants’ strategies of speech segmentation: clues from weak/strong words. Paper presented at the 20th Annual Boston University Conference on Language Acquisition, Boston, MA.

Johnson, N. F. (1970). The role of chunking and organization in the process of recall. In G. H. Bower (Ed.), The psychology of learning and motivation. Volume 4 (pp. 171-245). New York: Academic Press.

Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 3-11.

Jusczyk, P. W. (1997). The discovery of spoken language. Cambridge, MA: The MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants’ detection of sound patterns of words in fluent speech. Cognitive Psychology, 29, 1-23.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. Child Development, 64, 675-687.

Klatt, D. H. (1974). The duration of [s] in English words. Journal of Speech and Hearing Research, 17, 51-63.

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 Acoustical Society of America, 59, 1208-1221.

MacWhinney, B. (1983). Miniature linguistic systems as tests of the use of universal operating principles in second-language learning by children and adults. Journal of Psycholinguistic Research, 12, 467-478.

Mandel, D. R., Jusczyk, P. W., Kemler Nelson, D. G. (1994). Does sentential prosody help infants organize and remember speech information? Cognition, 53, 155-180.

Mandel, D. R., Kemler Nelson, D. G., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. Cognitive Development, 11, 181-196.

Meier, R. P., & Bower, G. H. (1986). Semantic reference and phrasal grouping in the acquisition of a miniature phrase structure language. Journal of Memory and Language, 25, 492-505.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Moeser, S. D., & Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. Journal of Verbal Learning and Verbal behavior, 11, 759-769.

Morgan, J. L. (1994). Converging measures of speech segmentation in preverbal infants. Infant Behavior and Development, 17, 389-403.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. Journal of Memory and Language, 35, 666-688.

Morgan, J. L., Meir, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language? Cognitive Psychology, 19, 498-550.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Development, 66, 911-936.

Newport, E.L., Gleitman, H., & Gleitman, L.R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. Ferguson & C.E. Snow (Eds.), Talking to children. Cambridge: Cambridge University Press.

Newsome, M., & Jusczyk, P. W. (1995). Do infants use stress as a cue for segmenting fluent speech? In D. MacLaughlin & S. McEwen (Eds.), 19th Annual Boston University Conference on Language Development (pp. 414-426). Somerville, MA: Cascadilla Press.

Nygaard, L. C., Sommers, M. S., Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. Perception & Psychophysics, 57, 989-1001.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. Journal of Acoustical Society of America, 54, 1235-1246.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? Journal of Experimental Psychology: General, 3, 264-275.

Peters, A. (1983). The units of language acquisition. Cambridge: Cambridge University Press.

Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. Journal of Child Language, 20, 43-60.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: a reevaluation. Journal of Experimental Psychology: General, 125, 123-138.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274, 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. Journal of Memory and Language, 35, 606-621.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 592-608.

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

Snow, C.E. (1972). Mothers' speech to children learning language. Child Development, 43, 549-565.

Swingley, D., & Pinto, J. P. (1997, April). The robust effect of one-year-olds' speech processing. Poster session presented at the biennial meeting of the Society of Research in Child Development, Washington, DC.

Tulving, E. (1962). Subjective organization in the free recall of "unrelated" words. Psychological Review, 69, 344-354.

Umeda, N. (1977). Consonant duration in American English. Journal of Acoustical Society of America, 58, 434-445.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: the role of market frequency. Journal of Memory and Language, 27, 71-86.

Valian, V., & Levitt, A. (1996). Prosody and adults' learning of syntactic structure.

Journal of Memory and Language, 35, 497-516.

Vihman, M. (1982). Formulas in first and second language acquisition. In L. Obler & L. Menn (Eds.), Exceptional language and linguistics (pp. 261-284). New-York: Academic Press.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. Language and Speech, 40, 47-62.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. Applied Linguistics, 16, 180-205.

Wong Fillmore, L. (1976). The second time around: Cognitive and social strategies in second language acquisition. Unpublished doctoral dissertation, Stanford University, California.

Zipf, G. (1949). Human behavior and the principle of least effort. New York: Addison-Wesley.

Author Note

Delphine Dahan, Department of Cognitive Science (now at the University of Rochester); Michael R. Brent, Department of Cognitive Science.

This work was supported by a grant from NIDCD (DC 03082) to Michael Brent, who is a member of the Center for Language and Speech Processing at Johns Hopkins University. We wish to thank the Max-Planck Institute for Psycholinguistics for technical help, Elizabeth Cole, Jamie Herslow, Kelly Lowe, Matthew Shomphe, and Marni Soupcoff, for their help in testing subjects, Nan Bernstein Ratner for giving us space and helping us recruit subjects at the University of Maryland, at College Park, Gerry Altmann, Anne Cutler, Peter Jusczyk, and Thierry Nazzi for helpful comments on earlier versions of the manuscript. Special thanks are due to Anne Cutler for inviting both of us to the Max-Planck Institute for Psycholinguistics (Nijmegen, The Netherlands) and letting us run pilot experiments there.

Correspondence concerning this article should be addressed to Delphine Dahan, Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, NY 14627, dahan@bcs.rochester.edu, or to Michael Brent, Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218, brent@jhu.edu.

Footnotes

[1] A _strong syllable_ in Cutler's sense is one with an unreduced vowel, even if it is not stressed (e.g., Fear, Cutler, & Butterfield, 1995). Although there is a strong correlation between unreduced vowels and stress in English, it is by no means a perfect correlation. Research on infants, however, has contrasted syllables that have both stress and an unreduced vowel with those that have neither.

[2] Within the model, the "relative frequency" of a word is actually computed as (f +1)/(m+1), where f is the number of times the word has occurred so far, and m is the total number of occurrences so far of all words. Thus, the relative frequency of a novel word, which has never occurred before, is 1/(m+1), not zero.

[3] The relative frequency of a novel unit is 1/(m+1), where m is the total number of word tokens encountered so far. Thus, if an utterance is treated as a single novel unit, the product of relative frequencies is simply 1/(m+1). If an utterance is divided into a familiar unit (e.g., _look_) and a novel unit (e.g., _here_) then the product of relative frequencies is the relative frequency of the familiar unit, (_freq_(look)+1)/(m+1), times the relative frequency of the novel unit, 1/(m+1). Since relative frequencies are always less than one, (_freq_(look)+1)/(m+1) × 1/(m+1) < 1/(m+1).

[4] Differences between test items located at the right and left edges of the utterance might be attributable to the fact that only those on the left receive initial stress in the long utterance, but an overall effect of word versus nonword condition could not be attributed to stress.

[5] This task was favored against a two-alternative forced-choice test, as used by Saffran, Newport, and Aslin (1996), where each word was paired with each nonword. In such a task, subjects may learn from the test itself, and may thus entertain hypotheses that they would not have entertained otherwise. For instance, for each pair, subjects would know that one of the items

is a word while the other is not a word (see Valian & Coulson, 1988, for similar argument). Moreover, the multiple presentations of a nonword, paired with all the different possible words, might cause the nonword to become more and more familiar, and therefore influence responses.

6.At the suggestion of an anonymous referee, we also performed a matched-pairs t-test comparing subjects' accuracy at responding to items in the word condition to their accuracy at responding to items in the nonword condition. This test also showed a significant difference ($t(47)=2.182$, $p<.05$, two-tailed). We also compared the percentage of yes responses in each condition to 50% and found that the difference was significant in the word condition ($t(47)=2.58$, $p<.01$) but not in the nonword condition.

7.At the suggestion of an anonymous referee, a 2×2 (word/nonword×left/right) ANOVA with the percentage of yes responses (0% or 100%) as the dependent variable was conducted. Like the Chi-square analysis, it did not show an interaction ($F(1,47)=0.855$). However, this ANOVA must be interpreted cautiously because each subject contributes either zero or one yes response in each cell, making this an extreme case of discrete data.

Table 1: Experiment 1. Materials organized into four item sets (presented with their letter code and phonemic notation) and two practice sets.

	Long utterance		Short utterance or Test item			
Item sets	abcde	/koʃedifenu/	ab	/koʃe/	cde	/difenu/
			abc	/koʃedi/	de	/fenu/
	fghij	/dobuneripo/	fg	/dobu/	hij	/neripo/
			fgh	/dobune/	ij	/ripo/
	klmno	/tegivemofu/	kl	/tegi/	mno	/vemofu/
			klm	/tegive/	no	/mofu/
	pqrst	/belizekufo/	pq	/beli/	rst	/zekufo/
			pqr	/belize/	st	/kufo/
Practice sets		/muzibo/	/bo/		/muzi/	
		/kezoru/	/ke/		/ru/	

Table 2: Experiment 1. Presentation of the design, limited to the item set abcde (the segmentation point is schematized by an underscore)

Group	Short utterance	Long utterance with predicted segmentation	test item	prediction
Group 1a	abc	abc_de	de	word
1b	abc	abc_de	cde	nonword
Group 2a	ab	ab_cde	de	nonword
2b	ab	ab_cde	cde	word
Group 3a	cde	ab_cde	abc	nonword
3b	cde	ab_cde	ab	word
Group 4a	de	abc_de	abc	word
4b	de	abc_de	ab	nonword

Table 3: Experiment 1. Subject classification as a function of the number of yes responses to test items in the nonword or word condition.

	nonword	word
0 yes response	16	4
1 yes response	19	25
2 yes responses	12	16
	47	45

Table 4: Experiment 1. Subject classification as a function of the number of yes responses to test items in the nonword or word condition, at the left or right edge of the long utterance.

	Test items on the left		Test items on the right	
	nonword	word	nonword	word
0 yes response	25	17	27	16
1 yes response	23	28	20	31
	48	45	47	47

Table 5: Experiment 1. Subject classification as a function of the number of correct responses to test items at the left or right edge of the long utterance.

	Items on the left	Items on the right
0 correct response	10	6
1 correct response	19	25
2 correct responses	16	16
	45	47

Table 6: Experiment 1. Subject classification as a function of the number of yes responses to 2- or 3-syllable test items in the nonword or word condition.

	Two-syllable items		Three-syllable items	
	nonword	word	nonword	word
0 yes response	29	15	23	18
1 yes response	19	32	24	27
	48	47	47	45

Table 7: Experiment 1. Subject classification as a function of the number of correct responses to two-syllable and three-syllable test items

	Two-syllable items	Three-syllable items
0 correct response	8	8
1 correct response	27	24
2 correct responses	12	12
	47	44

Table 8: Experiment 2. Mean accuracy and reaction time (in ms) to test items, classified by number of syllables (3- vs. 2-syllable items) and condition (nonword vs. word).

	Nonword Condition	Word Condition
3 syllable	73%	88%
	1715ms	1522ms
2 syllable	77%	85%
	1448ms	1454ms

Table 9: Experiment 3. Mean accuracy and reaction time (in ms) to test items, classified by number of syllables (3- vs. 2-syllable items) and condition (nonword vs. word).

	Nonword Condition	Word Condition
3 syllable	94%	92%
	1773ms	1736ms
2 syllable	54%	60%
	1425ms	1669ms

Table 10: Experiment 4. Mean accuracy and reaction time (in ms) to inference and control test items, classified by number of syllables (3- vs.2-syllable items) and condition (nonword vs. word).

	Inference test items		Control test items	
	Nonword	Word	Nonword	Word
3 syllable	78%	87%	89%	98%
	1909ms	1658ms	1722ms	1490ms
2 syllable	41%	59%	87%	98%
	1566ms	1546ms	1369ms	1304ms