Bootstrapping lexical acquisition: The role of prosodic structure[1]

ANNE CHRISTOPHE and EMMANUEL DUPOUX

Abstract

Speech runs continuously, without pauses between individual words. As yet there is little firm evidence that the speech signal contains usable information as to the location of word boundaries, therefore current psycholinguistic models of lexical access are based upon this assumption and rely solely on phonemic information (that is, they consider speech as an uninterrupted string of phones). In these models, segmentation into words is a by-product of lexical identification. Babies, who do not possess a lexicon to start with, have to use independent mechanisms in order to first build their input lexicons (a list of word-forms). After reviewing three potential sources of information (distributional regularity, phonotactics, and lexical biases) that may allow infants to start acquiring an input lexicon — to bootstrap *lexical acquisition — we propose that prosodic cues may be used by infants, and by adults, in order to segment the speech stream in prosodic units smaller than sentences, but bigger than words. Lexical acquisition, as well as lexical access, would be performed on the basis of this prosodically segmented pre-lexical representation. In addition, such a representation would be useful for the acquisition of phonology and syntax. In this paper, we present some experimental evidence that favors the existence of a prosodic segmentation strategy, using both adult and infant subjects. In addition, we discuss the implications of the hypothesis for models of speech processing — lexical access in adults, as well as lexical acquisition by children.*

1. The preparation of this paper was supported by a grant from the Direction des Recherches, Etudes et Techniques to the first author (nº 8780844). The Human Frontiers Scientific Program, as well as the Human Capital and Mobility Project, also supported the reported research. We would like to thank Jaques Mehler, John Morton, Geoff Hall, Brit van 0oyen and an anonymous reviewer for useful comments on the manuscript. Address for correspondence: MRC Cognitive Development Unit, 4 Taviton Street, London WC1H 0BT, UK.

The Linguistic Review 13 (1996), 383–412

0167–6318/96/0013–383

1. Segmenting speech into words

Identifying words in sentences is a necessary step in continuous speech processing. We compute the meaning of sentences by using the meaning of the words they contain in combination with a syntactic analysis. But even a superficial observation of the speech signal shows that spoken words are not separated from one another by silent pauses. Figure 1 shows the acoustic waves of the italicized part of the following two French sentences: *C'était son chat grincheux qui le rendait nerveux.* 'His grumpy cat made him nervous.'; *C'était son chagrin fou qui le rendait odieux.* 'His mad sorrow made him odious.'

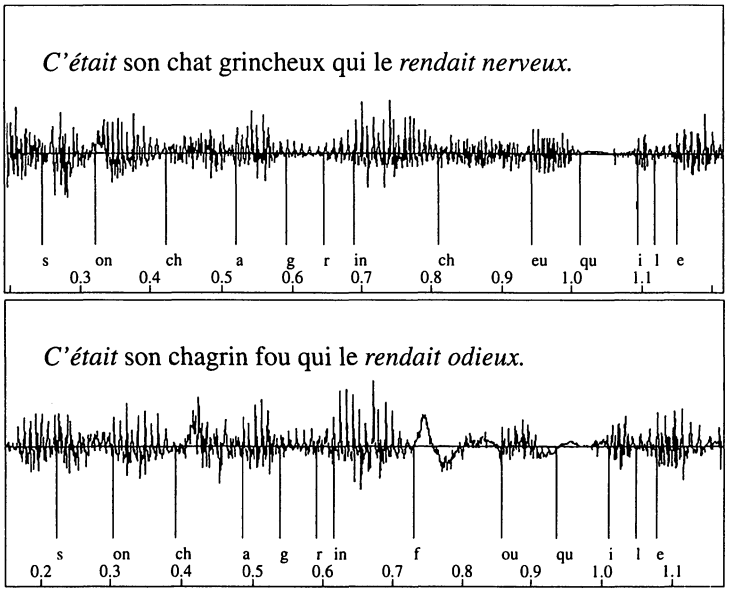

Figure 1. Illustration of the absence of pauses between words
Key: Acoustic wave for two French sentences (the beginning of each phoneme is marked by a vertical line; the scale is in seconds).

Note that the two sentences are identical up to the syllable "grin", as far as the sequence of phonemes is concerned, but they differ in the location of one word boundary: We see however that there is no pause between the syllables "chat" and "grin" in the first sentence, relative to the second one. Thus, at first sight, the speech signal resembles text written without spaces between words.

This fact has been taken at face value by psycholinguists trying to develop models of continuous speech processing. Here are a few examples taken from the psycholinguistic literature: "Examination of sound spectrographs ... show that words in sentences most often flow into one another without any intervening pauses." (Hayes and Clark 1970: 221); "it is extremely difficult to determine

strictly by simple physical criteria where one word ends and another begins in fluent speech" (Pisoni and Luce 1986: 4); "Spontaneous, continuous speech contains very few completely reliable cues to the location of word boundaries." (Shillcock 1990: 24); "... pauses between words are rare, and, indeed, no single consistent set of cues to word boundaries are present." (Echols 1993: 245). It is noticeable that the older assertions stress the absence of pauses between words (very correctly) while the more recent ones generalize to the absence of segmentation cues whatsoever. The crucial word here is "reliable" or "consistent": Indeed, research has shown that there is no such thing as a "boundary phoneme", that would be present at all and only at all word boundaries. Prosodic boundary cues, if they exist, are bound to be quantitative variations in the prosodic characteristics of individual phonemes, not qualitatively different events. In fact psycholinguists have generally assumed that prosodic information is extremely unreliable in that it is non-categorical and rests on small variations. Moreover, they have also assumed that even if prosodic information existed, it would not be necessary for the purpose of word recognition in adults. Indeed, adults know the words of their language and can therefore identify words without knowing their boundaries, just as we are able to read a text without spaces. Several models have been developed to offer mechanisms whereby lexical access can be completed from an uninterrupted and unstructured string of phonemes.[2] These models bypass the absence of boundary cues by directly identifying the words stored in the lexicon from the string of phonemes.

In this paper, we wish to raise two issues regarding these phoneme-centered, adult-centered models. The first one relates to acquisition. Even though the lexicon is available to adults this is not true of newborns. Thus, a non-lexical method of finding word boundaries must exist, otherwise infants would be unable to acquire the words of their language. More generally, we agree with Mehler, Dupoux and Segui (1990) that models of adults can be usefully constrained by considering how the acquisition process might take place. In this paper, we favor the view that the same kind of processing architecture underlies both word acquisition and lexical access, and hence that models of word segmentation relying solely on the lexicon are inadequate. The second assertion that we wish to challenge is that prosodic information is inexistent or not used by adults. Following Mehler, Bertoncini, Dupoux and Pallier (1994); Mehler, Dupoux, Nazzi and Dehaene-Lambertz (1996) and Cutler and Mehler (1993) we believe that non-phonemic information, for example rhythm or prosody, play a far more important role than has been traditionally recognized in psycholinguistics, both in speech acquisition and processing.

---

2. In the recent version of the Cohort model, there is no phonemic level, but direct access from features to lexical entries (Marslen-Wilson and Warren 1994; Lahiri and Marslen-Wilsos 1991); this makes no difference in the present discussion, where we wonder whether larger units, bigger than words, are used in the process of speech perception.

In what follows, we first review adult models of lexical access that only use an unsegmented string of phonemes as input. We then consider the segmentation issue from the perspective of acquisition: How can a lexicon be constructed when the input to the child is unsegmented?[3] Next, we offer a new model of speech perception based on the idea that prosody is used by infants (and by adults) in order to perform a first-order segmentation of the speech stream. In the remainder of this paper, we present some experimental evidence in favor of the existence of such a prosodic segmentation strategy.

1.1. *Finding word boundaries with phonemes and a lexicon*

In models that work strictly on the basis of phonemic information, there is no preliminary word boundary extraction, but boundaries are recovered once words are identified. In such models, word segmentation is a by-product of word identification. Cole and Jakimik (1980) first proposed a concrete solution to the problem. They suggested that word identification proceeds strictly from left to right: The first word at the beginning of a sentence is identified; then, a new lexical search is initiated from the point where this word ends, and so on. This process implies that there is only one word candidate at each point. However, upon hearing the syllable *car*, one cannot tell, on the basis of phonemic information alone, whether it is the word *car*, or the first syllable of the word *carpet*. Luce (1986) calculated that in American English, only 40 percent of the words (taking frequency into account) could be uniquely identified before or at their last phoneme (see also McQueen and Cutler 1992; Frauenfelder 1991). Thus, identifying words sequentially from left to right gives rise to multiple false alarms.

An alternative proposal is embodied in the TRACE model of word recognition (McClelland and Elman 1986). This model relies on the notions of multiple activation and competition: All the words compatible with the phonemic information are activated (multiple activation), and the words that share one or several phonemes inhibit one another (lateral inhibition, or competition). These two processes ensure that each phoneme is attributed to one and only one word. Still, some strings of phonemes allow several segmentations (for example,

3. Lexical acquisition in this paper refers to constructing the repertoire of word forms (phonetic or phonological representations) used when perceiving speech. Indeed, the lexical representations used for producing speech are acquired from the input that the child receives. Consequently, the production or output lexicon has to be somehow constructed from the perception or input lexicon, and thus the acquisition problem is most strongly posed for the input lexicon. Of course there is far more to lexical acquisition than building an input lexicon. For instance, how children discover the meaning of words is a lively topic of research. But building an input lexicon is a necessary first step.

succeed versus *suck seed*). In these cases, phonemic information will not suffice to resolve ambiguities; more abstract processing levels must be involved in the disambiguation, such as syntactic or semantic (or even pragmatic).

If the proportion of sentences with ambiguous segmentation were low, relying upon more abstract levels in the selection of possible segmentations would be of little consequence. In contrast, if the proportion is high, and particularly if the number of potential candidates per sentence is high, ambiguity implies an important mass of computations. Harrington and Johnstone (1987) calculated, for English sentences, the number of exhaustive segmentations (that is, strings of words such that each phoneme in the sentence is attributed to one and only one word), and they observed that less than 5 percent of sentences allowed a unique segmentation when syntactic constraints were not allowed to operate. It therefore seems that when speech is considered as a string of phonemes, full knowledge of the lexicon is not sufficient for identifying words: An important number of sentences require, for the mere identification of the words that compose them, the intervention of abstract levels of processing such as syntax and semantics, in a way that remains to be specified.

In an attempt to reduce this problem, Anne Cutler and her colleagues developed a search heuristic for the case of English, that is, a way to reduce the number of possibilities that are examined in vain (see, for example, Cutler 1990). This proposal rests on the fact that in English, more than 90 percent of content words (that is, nouns, verbs, adjectives, but not grammatical words such as articles) begin with a strong syllable (Cutler and Carter 1987). Strong syllables contain a full vowel, in contrast to weak syllables that contain a reduced vowel. Thus, for English listeners, it is advantageous to hypothesize a word boundary before each strong syllable.[4] Anne Cutler and her colleagues found experimental evidence that adult English listeners do indeed make use of this characteristic of English, by using among others a task of word detection in non-words, or word-spotting (Cutler and Norris 1988; McQueen, Norris and Cutler 1994) and by analyzing perceptual errors (Cutler and Butterfield 1992). This strategy has been offered for the treatment of English, and so far it seems to be specific to this language. It may be that it will generalize to other languages that embody the strong–weak syllable distinction (such as Dutch, in which

4. This strategy is also prosodic in that it amounts to say that listeners posit a word boundary at foot boundaries, and the foot is a constituent of the prosodic hierarchy. But it differs from the "prosodic segmentation hypothesis" advocated in this paper in two respects: First, the strong–weak distinction may be seen as phonemic rather than prosodic in nature, given the fact that all the weak syllables and only them have a schwa as nucleus. Second, the foot is smaller than the word in the prosodic hierarchy, which means that some foot boundaries occur within words, thus creating false alarms. In contrast, marking of prosodic units bigger than the word lies in quantitative variations in the duration, pitch and energy of individual phones (which may be less reliable than a phonemic difference such as vowel quality) and there are no false alarms, although presumably not all word boundaries will be marked.

the same predominance of strong syllables at the beginning of content words is found, see Quené 1992). However, other ways to solve the problem for languages where this distinction does not apply remain to be discovered (for example, French, Spanish, or Japanese, none of which possess a distinction between reduced and full vowels).

Some studies were aimed at experimentally evaluating the multiple activation plus competition account of lexical access. McQueen *et al.* (1994) and Norris, McQueen and Cutler (1995) focused on the competition component of the strategy, and used the word-spotting task mentioned above. The authors observed significant effects of competition, both when only one long word competed with the target (for example, *mess* was more readily detected in *nomes* than in *domes* which forms the beginning of *domestic*) and when several words competed for one phoneme of the target (for example, *mint* was more readily detected in *mintaup* than *stamp* in *stampidge*, because few words begin with *tau* as in *town*, whereas many begin with *pi*). These effects have been implemented in the Shortlist model (this model also rests on multiple activation and competition for lexical segmentation, but is implemented in a psychologically more realistic fashion than the TRACE model, see Norris 1994).

Other studies focused on the multiple activation part of the strategy, using the cross-modal priming technique (Swinney 1981; Zwitserlood 1989).[5] Three recent studies pitted multiple activation against acoustic cues, using naturally produced sentences, the segmentation of which was locally ambiguous (but potentially containing prosodic boundary cues). Tabossi (1993) showed that an associate of the word *visite* was primed in both of the following Italian sentences: ... *visite di* altri membri ... (... *visits of other members* ...) and ... *visi tedi*ate e stanchi (... *faces bored and tired* ...). This result suggests that acoustic information about the presence of a word boundary between *visi* and *te* in the second sentence is either not present or not used to block access to *visite*. Similarly, Shillcock (1990) showed that an associate of *bone* was primed when subjects listened to a sentence containing the word *trombone*. This may suggest that every embedded word and every word resulting from the concatenation of adjacent word fragments gets activated during word recognition, a result that we find somewhat puzzling. However, Gow and Gordon (1995) failed to observe priming of an associate of *lips* in a sentence containing *tulips*, in contrast with

5. This technique was first used by Swinney, Onifer, Prather and Hirshkowitz (1979) to study the processing of ambiguous words in sentences. It consists of visually presenting a word that is sometimes semantically associated to an auditorily presented word and sometimes not. An advantage in reaction time for identifying associated words over non-associated words is called *a priming effect.* The classical interpretation of this effect is that subjects respond faster to the associate because the auditorily presented word activated all its associates; thus, when one of these words is later visually presented, it is processed faster, and thus responded to faster than the non-associate.

the above-mentioned result (although they did observe, just as Tabossi 1995, priming of an associate of *tulips* by a sentence containing *two lips*). It seems that more research is needed to clarify this issue — in particular, comparison between studies should include a comparison of the kind of word boundaries studied.[6] To date it seems that most results are compatible with the hypothesis that adults tend to activate all potential candidates, irrespective of acoustic boundary cues.

In brief, multiple activation and competition offers a precise mechanism by which adults could recover words from an uninterrupted string of phonemes, provided it is supplemented with semantic and pragmatic information to help with cases where segmentation is ambiguous. Particularly interesting additions are statistical biases such as that proposed by Anne Cutler and her colleagues for English in that they are likely to reduce the need for higher level information. An exact evaluation of the performance of such competition plus biases mechanisms still needs to be carried out on natural corporae. However, there is already experimental evidence that adult listeners use these mechanisms, although the interaction between these processes and prosodic boundary cues remains unclear.

Yet, whether this line of research is correct or not, it cannot be the whole story. Indeed, infants initially lack all of the lexical (not to speak of the pragmatic) information that is supposed to make word segmentation possible. According to such models, how could then babies acquire a lexicon? One possible answer is that babies use some other mechanism and that these issues should not bear on the adequacy of adult models. On the contrary, we take the view that adult and infant processing systems are intimately related, one being a matured and more specialized version of the other. If this is true, then surely acquisition issues will have an impact on adult models. In the next section, we examine how infants start constructing an input lexicon from an uninterrupted string of phonemes.

1.2. *Finding word boundaries with phonemes and no lexicon*

In this section we examine potential solutions to the word boundary problem. We set two criteria for what we will consider a potential solution to the problem. Firstly, it must be explicit, that is, only mention processes that are sufficiently detailed that an implementation may be envisioned. Secondly, the

---

6. For instance, Gow and Gordon's (1995) study included a wide variety of cases. The smallest boundary studied was a word boundary within a clitic group, such as *a claim* versus *acclaim*, while the biggest "word boundary" studied actually coincided with an intonational phrase boundary, such as: *When the first runners [pass tell] them their times* versus *When the first runner's pastel shorts came into view someone made a crack.*

potential solution must be reasonably complete and accurate, that is, it must discover a reasonable number of word boundaries and must not posit too many irrelevant ones.

Discovering word boundaries through word identification is of no use in constructing an input lexicon. Yet, this fact alone is not sufficient to reject phoneme-based models. It may be that language is structured in such a way that statistical information concerning the way phonemes are distributed is sufficient to retrieve word boundaries. If this were the case, infants might compute some of these statistics and hence extract word boundaries. As formulated by Brent, Cartwright and Gafos (in press), "distributional regularity refers to the intuition that sound sequences that occur frequently and in a variety of contexts are better candidates for the lexicon than those that occur rarely and in few contexts." Note that this intuition is not very explicit, however Brent *et al.* (in press) formalized it and offered algorithms to test the feasibility of such a proposal. They showed, using a transcript of speech directed to children, that an algorithm based on distributional regularity can discover about 40 percent of the words in their right place with an accuracy of 47 percent, a performance significantly better than that of an algorithm which places word boundaries at random. This algorithm shows that distributional regularity brings useful information and hence might be a component to the full solution. Is there any evidence that infants pay attention to distributional regularities? Goodsitt, Morgan and Kuhl (1993) devised an ingenious experimental test to asses infants' sensitivity to co-occurrences between syllables. They showed that eight-months-old infants were better at processing trisyllabic strings where a pair of syllables consistently co-occurred (such as *gakoti* and *tigako* which may be represented with only two units, *gako* and *ti*,) than trisyllabic strings where all syllables behaved independently (such as *gakoti* and *tikoga,* which have to be represented with three independent units, *ga*, *ko* and *ti*).

Another intuition that could lead to improved algorithms comes from the statistical distribution of sounds within words (as opposed to sounds within utterances). We will explore two variants of this intuition: Phonotactic constraints and lexical statistics.

Phonotactic constraints refer to the fact that sounds do not occur freely within words. Certain sounds (or combination of sounds) only occur at word edges, others only occur word internally. English phonotactic constraints are mostly expressed in terms of the consonant clusters that may appear at the beginning and end of syllables; thus, a string of phonemes as /dstr/ necessarily contains a word boundary between the /d/ and the /s/ (as in *bad string,* for instance) because it is not a possible word-internal cluster, and all other segmentations give rise to illegal word beginnings or endings. Other examples of phonotactic constraints include the vowel harmony rules that operate in languages such as Turkish (there is a high probability of encountering a word boundary between

two vowels that do not share a given feature). It can be shown that by using constraints on the consonant clusters of English one can improve the performance of algorithms which try to locate words without a lexicon (Brent *et al.*, in press; Cairns, Shillcock, Chater and Levy, in press). Given that not all word boundaries are marked phonotactically, it is not surprising that performance is still far from perfect. With respect to babies, Friederici and Wessels (1993) showed that nine-month-old Dutch infants listened longer to syllables that respected the phonotactic constraints of their language (namely, syllables that had possible onset and coda consonant clusters such as *BRef* and *muRT* versus syllables that had impossible clusters, such as *feBR* and *RTum*). It is therefore conceivable that infants make use of the phonotactic constraints of their native language in order to help word segmentation from the age of nine months.

At the margins of the phonological regularities of a language's lexicon there are regularities that are not all or none, but probabilistic. For instance, English content words predominantly start with syllables with a full vowel (Cutler 1996). Hence, it may be useful in English to posit or favor a word boundary before a strong syllable. To our knowledge, however, such probabilistic regularities have not been implemented in learning algorithms although they are likely to improve performance (but also likely to be incomplete, since words starting with a reduced vowel will not be learned, or be misparsed). In fact, Jusczyk, Cutler, and Redanz (1993a) showed that nine-month-old American infants chose to listen longer to lists of bisyllabic words that exemplify the most common pattern in English, namely strong–weak words, rather than weak-strong words. Moreover, very recent experimental work with American infants from eight to twelve months of age suggests that they actually use this regularity of English when hypothesizing words from continuous speech passages (Newsome and Jusczyk 1995).

In brief, we discussed three potential sources of information, namely distributional regularity, phonotactics, and lexical statistics, that have been shown to be useful for finding word boundaries, and have some empirical support in the behavior of infants. However, none of these mechanisms offers a complete solution to the word segmentation problem. There is a further and more damaging problem for the last two sources of information: Both phonotactic constraints and lexical statistics are largely language-specific and hence have to be acquired beforehand. However, if a lexicon is needed in order to extract phonotactics or lexical statistics, then we are back to square one. You need the lexicon in order to learn the lexicon.[7] To break this vicious circle, it

7. In fact, it looks like the infants do not need to have a lexicon in order to know the properties of the words of their language. As we have seen, language-specific phonotactics (Dutch infants) and the English strong–weak lexical bias (American infants) emerges before the age of nine months, that is, before infants show any knowledge of specific words. How this may be achieved is the purpose of the rest of this paper.

could be hypothesized that language-specific regularities are acquired on the basis of certain available boundaries. Fortunately, continuous speech is interrupted by pauses from time to time, if only because speakers have to take in breath. Whether these pauses are sufficient to bootstrap acquisition is an open issue, but at least one can see the usefulness for postulating *some* boundary information in addition to segmental information. In the next section, we explore the possibility that boundary information going beyond pauses, that is, prosodically marked boundaries may be used in order to bootstrap lexical acquisition.

1.3. *Finding word boundaries with no phonemes and no lexicon: The prosodic segmentation hypothesis*

Our proposal is that prosody may help by providing a first-order segmentation of the speech signal. Further segmentation would involve the processes outlined above, which would take place within prosodic units smaller than whole utterances. Prosodic boundaries would allow the acquisition of language-specific knowledge linked to boundaries (such as phonotactics). Before we go any further, let us first comment on what we mean by prosody. Prosodic constituents can be defined in at least two different ways: Phonological and psychological.

Recent developments in phonological theory showed the necessity of an independent prosodic hierarchical structure to mediate between the morpho-syntactic hierarchy and phonological phenomena (Selkirk 1984; Nespor and Vogel 1986). The constituents of this prosodic hierarchy are defined as the domains of phonological rules and are derived from the morpho-syntactic hierarchy (in this paper, we will refer to the prosodic hierarchy as defined by Nespor and Vogel 1986, namely the syllable, foot, phonological word, clitic group, phonological phrase, and intonational phrase). As phonology mediates between abstract representations and the actual production of speech, phonological constituents are correlated but do not always coincide with syntactic constituents.

Psychological prosodic constituents are used in the process of speech production and speech perception, and are therefore expected to be marked in the speech signal. With regard to the existence of prosodic boundary cues, a number of phonetic studies have found systematic influences of word boundaries on acoustic parameters such as the duration, pitch and energy of individual phonemes.[8] Thus, all languages that have been studied to date have shown evidence of a word-initial consonant lengthening (see, for English: Umeda 1977;

8. Since most older phonetic studies do not refer to prosodic constituents, but only to words, or at best syntactic constituents, it is hard to directly use their findings to speculate about the marking of prosodic constituents. Hence the continued use of "word boundary", even though it may cover many distinct cases.

for Dutch: Quené 1992; for Czech: Lehiste 1965; for Estonia: Lehiste 1966; and for Italian and Swedish: cited in Vaissière 1983). It is also quite common to observe word-final vowel lengthening (for English, see for example: Nakatani, O'Connor, and Aston 1981; for Finnish: Lehiste 1965; for Dutch, Quené 1992; for French, see for example: Rietveld 1980; and for Spanish and Japanese: Hoequist 1983). More to the point, Grosjean and his colleagues have proposed a structure called "performance structure" that could account for the distribution of these parameters (Grosjean, Grosjean and Lane 1979). Gee and Grosjean (1983) showed that the performance structures, uncovered by asking people to read very slowly and measuring pauses between words, correlated extremely well with the constituents established by prosodic phonology. In a recent study involving French, Monnin and Grosjean (1993) bypassed the need for artificially slow speech by measuring the length of the final vowel of each word in a set of sentences read at normal rate (including the length of a pause following the word when there was one). The constituents emerging from this set of measurements corresponded remarkably well to the constituents from the prosodic hierarchy.[9] However, the mapping is not perfect, and other factors play a role, most notably the syllabic length of the constituents. Thus, *un beau chat blanc* (*a nice white cat*) makes only one constituent, while *un spectaculaire chat indonésien* (*a spectacular Indonesian cat*) makes two — with a break after *chat* at the phonological phrase boundary and not before *chat* which would yield two units of equal syllabic length. We suggest that abstract constructs such as prosodic constituents indicate where boundaries may fall, while performance constraints such as the number of syllables within each unit determine which of the possible boundaries are actually marked (this is in fact the approach taken in some text-to-speech systems, see for example, Dirksen 1992, cited in de Pijper and Sanderman 1994; see also Delais 1995, for a proposal of competence and performance constraints interacting to give the prosodic units). We claim that these psychological prosodic units used in production, and marked in the speech signal, are recovered on-line by the hearer while perceiving speech. That is, we believe that prosodic constituents are perceived on the basis of prosodic information like duration, energy and pitch, and that they are linguistically relevant. What is the size of the units we are postulating? Although we are

9. Particularly striking is the fact that the boundary between a noun and an adjective is consistently more marked when the adjective follows the noun than when the adjective precedes the noun. The authors comment that no syntactic theory can account for this fact. However, prosodic phonology gives a ready explanation to this fact: Because phonological phrases are defined as including a lexical head and everything on its left (more generally, non-recursive side) up to another lexical head that is not included within its maximal projection, there is a phonological phrase boundary between a noun and the adjective that follows it (such as in *le chat # blanc – the white cat*) but not between a noun and the adjective that precedes it (such as in *le beau chat – the nice cat*).

leaving the precise nature of the units used in perception open to investigation, we believe that they are likely to be of a size smaller or equal to phonological phrases, and possibly clitic groups.

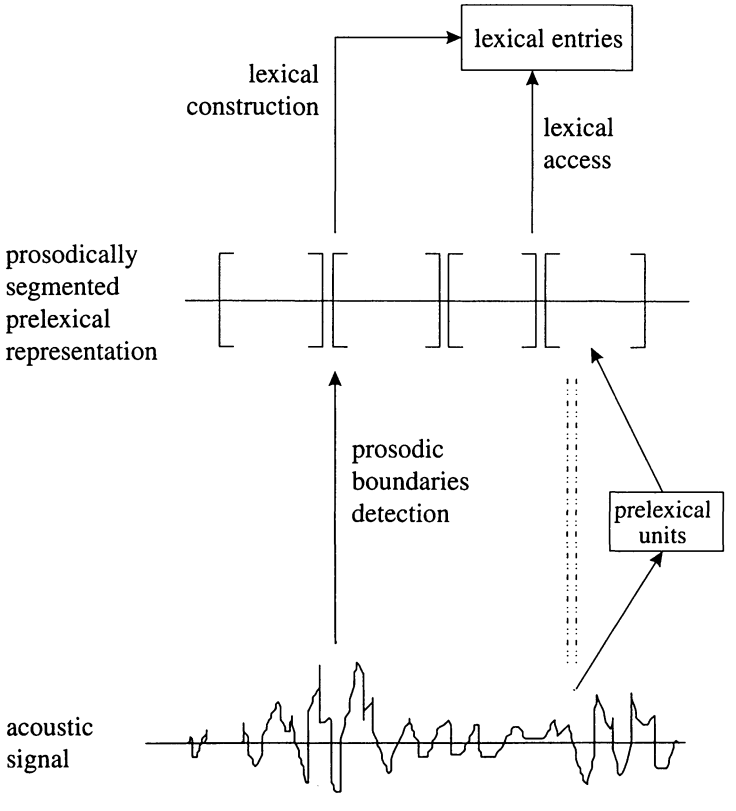

Figure 2. The prosodic segmentation hypothesis
Key: Its main feature is the existence of a prosodically segmented prelexical representation, which is used for lexical access by adults and for lexical acquisition by infants. Prosodic segmentation and prelexical processing are assumed to be independent.

Figure 2 illustrates the *prosodic segmentation* hypothesis. The main feature of the hypothesis is that there exists a prosodically segmented representation that is used for constructing lexical entries and for later lexical access in the adult. Of course, even if babies construct their lexicon on the basis of a segmented representation, it could be that adults abandon this information to only exploit phonemic information for lexical access. However, as we said in the introduction, we consider it reasonable to assume that infants and adults share the same

basic processing architecture. Therefore, lexical access in adults is assumed to operate from the same prosodically segmented representation.

Schematic as it is, this picture already makes clear certain predictions which follow from the hypothesis that speech is spontaneously perceived as a string of prosodic units. First, we expect the prosodic segmentation mechanism to be independent from phonemic processing: As a consequence, it should, to a certain extent, work on impoverished speech where prosody is preserved (such as low-pass filtered speech). Secondly, we expect other segmentation mechanisms (for example, through identification, phonotactics, lexical biases) to operate within prosodic units. Thus, we would expect no effect of lexical ambiguity or competition across a marked prosodic boundary. That is, in cases where a syllable can be attributed to the preceding or the following word we expect a parsing without cost whenever the boundary is prosodically marked.

Before we turn to experimental evidence, we wish to make a final comment. For the process of lexical acquisition, we hypothesized that knowledge of prosodic boundaries would reduce the workload of other processes (such as distributional regularities or phonotactics) in that these processes would apply within units smaller than utterances. Note that prosodic boundaries are qualitatively different from boundaries hypothesized by these other mechanisms in that they are linguistically relevant: As a consequence, whenever a boundary is marked, it will correspond to a word boundary, although the reverse is not true and many word boundaries will not be marked. We think that any serious error in the word form lexicon is unlikely to be corrected by later semantic/syntactic acquisition. On the contrary, we believe that if the lexicon is incorrect initially, this may induce a chain of acquisition problems in reference fixation and syntax acquisition. In this view, under-segmentation is preferable to over-segmentation. Prosodic segmentation is desirable in that it is a process that does not posit boundaries where there are none.

To summarize this section, the prosodic segmentation hypothesis is based on the idea that speech is spontaneously perceived as a string of prosodic constituents and that this perceptual process is one of the earliest stages of processing, which then feeds into the lexical level. On this view prosodic units would be used by babies to construct entries in their input lexicon, and by adults to access these entries. In what follows, we first examine the evidence available for the psychological reality of prosodic structure. We then turn to studies that bear on the use of boundary markers by adults in the on-line processing of continuous speech. Finally, we consider infants and review studies about the perceptibility and use of prosodic information in the first year of life.

2. Some evidence for the psychological reality of prosodic structure in perception

2.1. *Off-line evidence*

Some psychological studies have shown that subjects were able to find word boundaries better than chance even when the sequence of phonemes in itself was insufficient to decide where the word boundaries fell, such as in nonsense or ambiguous strings of syllables. Listeners relied on prosodic cues in such cases (see, for example, Nakatani and Schaffer 1978; Rietveld 1980).

Such studies prove that prosody may be used to disambiguate ambiguous strings of phonemes, but not that it encodes a prosodic hierarchy. Two recent studies have focused on whether listeners could exploit prosodic cues to retrieve prosodic constituents. De Pijper *et al.* (1994) studied how listeners rated the depth of all word boundaries in a set of sentences, either with normal speech or with speech that was modified in order to destroy phonemic information while keeping prosodic information intact (so-called *delexicalized speech*). They showed that subjects' ratings were extremely similar in the two conditions. This indicates that subjects were able to actually hear prosodic boundaries in the speech signal and that perceived depth was not the result of a reconstruction process involving subjects' knowledge of the phonology and syntax of their native language. They also showed that the boundaries receiving higher markings (the deepest ones) corresponded remarkably well with the boundaries placed by an algorithm that used both a morpho-syntactic analysis of the sentence in order to derive a two-level prosodic hierarchy (corresponding roughly to phonological and intonational phrases) and some performance constraints (Dirksen 1992).

Gussenhoven and Rietveld (1992) manipulated the duration of a phoneme just preceding a boundary (either a foot, word, phonological phrase, or intonational phrase boundary) and asked subjects to judge whether the duration of this phoneme was appropriate or not. This gave an indirect measure of how much pre-boundary lengthening was expected by subjects before each of the boundaries studied. Although the authors did find a trend for bigger boundaries to call for more lengthening (thus reflecting the hierarchical structure) this trend was significant only between the foot and the bigger boundaries (a foot boundary accepting less lengthening than either a word, phonological phrase, or intonational phrase boundary). To sum up, there is some evidence that prosodic constituents have a measurable influence on the speech signal, and that listeners are able to exploit these prosodic cues. It seems reasonable to think of these units as derived from the constituents of the prosodic hierarchy (itself derived from the morpho-syntactic hierarchy) while performance constraints such as syllabic length would determine which boundaries are actually marked in the signal. From the studies available, it seems that two levels of the hierarchy

could be accessible, corresponding to the intonational and phonological phrases, modulo performance constraints. We now turn to studies that bear on the use of boundary markers by adults in the on-line processing of continuous speech, as opposed to *off-line* processing, that takes place after the automatic processes of speech perception, and may involve meta-linguistic skills.

2.2. *Phoneme detection within words*

It may be the case that if we synthesized speech that contained only information about the phonemes and so prosodic information at all, adults would still be able to understand it, although perhaps with more effort, just as processing text written without spaces is more difficult than processing text with spaces (one should note that backtracking is hardly possible for speech, and that current research in speech synthesis aims at making it more intelligible by adding good prosody). The prosodic segmentation hypothesis claims that not only is prosody used, but also that it is used in the very first stages of speech processing. Lexical access is then performed on the units resulting from the prosodic segmentation. We contrast this hypothesis with current psycholinguistic models of lexical access in which lexical identification operates first, and other kinds of information, such as syntactic, semantic, and maybe also prosodic, intervene only afterwards, as a way of choosing between multiple solutions. If we asked subjects to perform a given experimental task after the whole process of sentence comprehension has been completed, we would most likely find that all the available sources of information have been used. Therefore, in order to distinguish between the two models we need to study speech processing as and when it happens (that is, on-line). We used the phoneme detection task to study the role of prosodic boundary information in speech processing. Phoneme detection (originally designed to study sentence processing, see Foss 1969) has been used in the past to study lexical access. In particular, several experiments have shown that subjects respond faster to the first phoneme of a word than to the first phoneme of a non-word (see, for example, Rubin, Turvey and Gelder 1976, and Dupoux, Christophe and Mehler, submitted, for a review). Different interpretations of this result have been offered. All invoke two levels of representation, pre-lexical and lexical. The pre-lexical level of representation may be constructed directly from the speech stream, and is available both for words and for non-words. The lexical level of representation contains phonological representations for the words that the subject knows. All interpretations attribute the faster responses to words than to non-words to the influence of the lexical level of representation. Christophe, Pallier, Dupoux and Mehler (submitted) have studied the activation of mono-syllabic words embedded in longer words. For instance, in French the first syllable of the word *bosquet* 'grove' is also a word, *bosse* 'bump'. In contrast, the first syllable of the word *mosquée* 'mosque' is not a word. If all the words compatible with the phonemes are

automatically activated during speech processing, then one would expect that upon hearing *bosquet*, subjects would momentarily activate the word *bosse*. Thus, one might expect faster responses to detect the target phoneme [s] in *bosquet* than in *mosque*, because the word *bosse* would accelerate reaction times in *bosquet*, while this would not happen for *mosquée*. The results show that although subjects respond faster to *bosse* than to *mosse*, thus replicating the lexical status effect for monosyllabic words, responses to *bosquet* and *mosque* are similar in speed (see figure 3 — a full report of the experimental materials and results is to be found in Christophe *et al.* submitted).

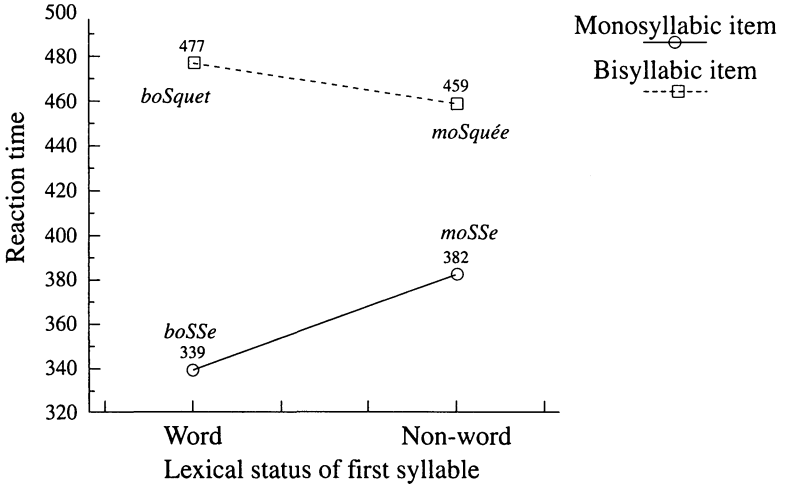

Figure 3. Mean reaction times for mono- and multisyllabic items which first syllable is or is not a word (adapted from Christophe *et al.* submitted)
Key: Subjects detected the phoneme is in *boSSe,* which is a word, faster than in *moSSe,* which is not a word. In contrast, responses to *s* in bosquet and *moSque* are similar in speed.

According to the prosodic segmentation model, *bosse* is not activated in *bosquet*, because there is no prosodic boundary between the two syllables of *bosquet*. Thus, the conditions for activating *bosse* are not all fulfilled. This experiment shows that when prosodic information is available (in this case a silent pause at the end of each word), it is used to constrain lexical searches. However we still need to establish whether prosodic boundary information is also used in continuous speech where prosodic boundaries are less well-marked than in lists of isolated words.

2.3. *Phoneme detection within sentences*

In addition to the lexical status effect for monosyllabic items, we also observed a *syllable length* effect in the above-mentioned experiment. That is, monosyllabic items are responded to faster than multisyllabic items. Although there are several possible interpretations for this effect (for instance, a pure length effect, or a lexical effect, see Christophe 1993), it tells us something about the access to the unit *word* — either the lexical representation of words, or the prosodic units corresponding to words, or both. In the case of the word lists that we have studied thus far, access to the prosodic unit corresponding to words was trivial since each word was surrounded by silence. However, if the syllabic length effect were to show up in sentences, this would imply that adult subjects have fast access to words even in continuous speech.

We tested this idea by using sentences, the segmentation of which is locally ambiguous. As mentioned above, lexical identification is probably one of the processes that allow adults to retrieve word boundaries. Therefore, if we want to assess specifically the role of prosodic information, we have to delay lexical processing to some extent (otherwise it is hard to tease apart which information, lexical or prosodic, is used first). The sentences used in this experiment were paired such that two sentences were identical up to a given syllable as far as the sequence of phones is concerned, but differed in the position of one word boundary (for example, *C'était son* chat grin*cheux* ... versus *C'était son* chagrin *fou* ...). In terms of the prosodic hierarchy as defined by Nespor and Vogel (1986), this word boundary is a clitic-group boundary; it may or may not be a phonological phrase boundary as well depending on whether restructuring occurred or not.[10] If continuous speech is treated at first like a string of phones, then one would expect the two sentences to be treated at exactly the same speed (since the boundary would be discovered only after identification of the words, that is, after processing of the third syllable in the string, and presumably after the response to the initial phoneme has been made). In contrast if a perceptually salient prosodic boundary between the syllables *cha* and *grin* is contained in the first sentence but not the second one, subjects should respond faster to the first phoneme of *chat* than to the first phoneme of *chagrin*, just like they do in lists of isolated words.

The results showed that subjects responded to monosyllabic words faster than to multisyllabic words (see figure 4; again, a full report of the experimental materials and results is to be found in Christophe *et al.* submitted). Thus, the

10. Restructuring refers to the collapsing of two adjacent phonological phrases: The phonological conditions for restructuring were met in the sentences used in these experiments (the second phonological phrase should be non-branching and should be a complement of the first one); furthermore, both phonological phrases concerned were rather short (at most six syllables altogether). Therefore, it seems probable that the boundary studied was not a major one.

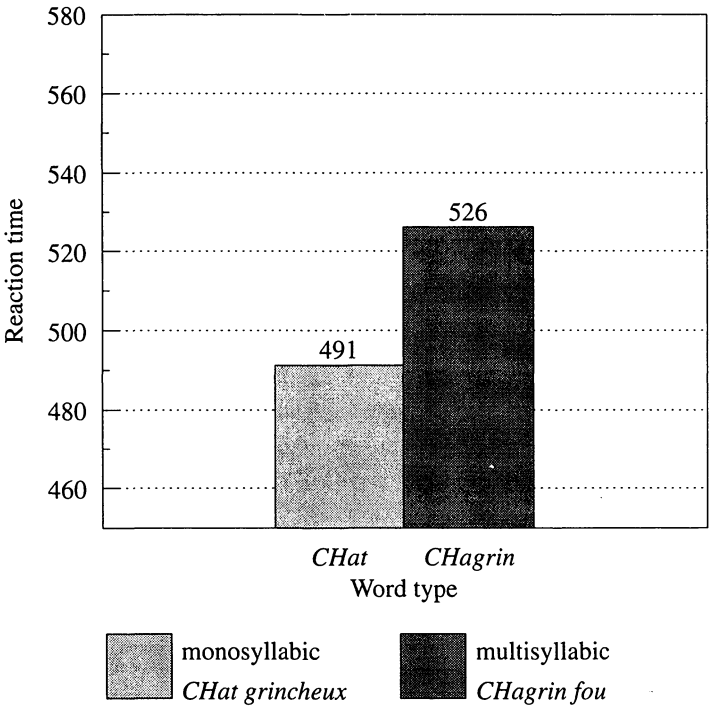

Figure 4. Mean reaction times for mono- and multisyllabic items in sentences whose segmentation is ambiguous, such as "chat grincheux" versus "chagrin fou" (adapted from Christophe *et al.* 1995)

syllabic length effect that was first observed in lists of words, was replicated in sentences whose segmentation is locally ambiguous. This is the outcome that was predicted by the prosodic segmentation hypothesis.

In the same experiment, we found that reaction times to word-initial phonemes were significantly faster than reaction times to word-medial phonemes. This may indicate that subjects find it comparatively easier to focus their attention on the beginning of words (or that word-initial phonemes are more clearly produced). This, in turn, may imply that subjects know where words begin at an early stage of speech processing. This interpretation is very easy to test experimentally: All that is required is a slight modification of the experimental task and ask subjects to respond only when the target phoneme is word-initial. If the effect of position in the word is an artifact (due for instance to word-initial phonemes being clearer) this modified task should be more difficult for subjects, since they not only have to recognize the target phoneme, but also

to check that it appears at the beginning of a word. In contrast, if the position effect reflects a natural tendency for subjects to focus their attention on the beginning of words, then they should be as fast at specifically detecting word-initial phonemes as at detecting phonemes anywhere in a word. Thus, using exactly the same sentences as in the previous experiment, we instructed subjects to only respond to word-initial phonemes. We found that subjects found it very easy to detect word-initial phonemes: They did not make more errors, nor responded more slowly than when they had to respond to phonemes in any position. The comparison between these two sets of instructions suggests that subjects have a spontaneous and easy access to the beginnings of words (at least for the words that were used in this experiment, that is, nouns; the effect remains to be tested on other types of words). To summarize the results so far, we observed a syllabic length effect, with monosyllabic words being responded to faster than multisyllabic words, in conditions where lexical access on the basis of phonemic information alone required the processing of the third syllable after the target phoneme. We suggested that the most plausible interpretation of this effect is that the boundary between the noun and the adjective is marked prosodically. As we mentioned above, this boundary is at least a clitic group boundary, and possibly a phonological phrase boundary as well. Moreover, we observed an advantage for word-initial phoneme processing: Subjects responded faster to these phonemes than to word-medial ones in a generalized phoneme detection task, and were able to respond to them selectively, without any additional cost, in a word-initial phoneme detection task. All of these results indicate an on-line exploitation of prosodic information during sentence processing, at least as measured by this particular experimental task, and support the prosody-based model of speech processing that we offered earlier.

3. Use of prosodic information in acquisition: Infant studies

The main reason that led us to investigate the perceptibility and use of prosodic boundary cues, is the acquisition problem. We suggested that knowledge of certain prosodic constituents' boundaries would be extremely useful to infants for acquiring a lexicon (a repertoire of word forms). The available prosodic boundaries would also allow other non-lexical segmentation mechanisms (such as phonotactics) to be bootstrapped.

3.1. *Perception of prosodic units by babies*

Peter Jusczyk and his colleagues have conducted a series of experiments to investigate which type of unit infants perceive in continuous speech. They presented infants with continuous speech that was interrupted by one-second pauses. For each unit tested, pauses were placed either between two units or

within a unit. The experimental measure was the amount of time babies chose to keep listening to each of the two versions (in this kind of experiment, stimuli are typically presented through loudspeakers that are situated on each side of the infant; the sound continues as long as the infant looks at the loudspeaker, and stops when the infant stops looking).

Results from American infants of nine, six, and even four-and-a-half months show a preference for stimuli that are interrupted at syntactic clause boundaries (or intonational phrase boundaries), relative to stimuli that are interrupted within a clause (Hirsh-Pasek, Nelson, Jusczyk, Cassidy, Druss and Kennedy 1987). This result indicates that these boundaries are marked, and that infants are sensitive to the corresponding cues from the age of four-and-half months. The authors report a decrease in pitch at the end of the clause, as well as a lengthening of the final syllables. These characteristics are usually considered to be universal (Bolinger 1978; Cruttenden 1986). Accordingly, when American four-and-half-month-olds were presented with Polish sentences, results showed that they also preferred stimuli interrupted at clause boundaries (Jusczyk, Kemler–Nelson, Hirsh–Pasek and Shomberg, in preparation). In all of these studies, the results were identical when stimuli were low-pass filtered, so as to leave only prosodic information in the signal. Jusczyk and Krumhansl (1993) obtained similar results with musical phrases. These results could be interpreted as a preference of the general perceptual system (not necessarily linguistic) for pauses preceded by a decrease in rate of speech and in pitch.

Would this technique enable us to evaluate infants' sensitivity to units smaller than clauses? Gerken, Jusczyk and Mandel (1994) reinterpreted Jusczyk, Kemler-Nelson, Hirsh-Pasek *et al.* (1992) results in terms of prosodic, rather than syntactic units. Their results suggest sensitivity to boundaries smaller than intonational phrases boundaries, corresponding to phonological phrase boundaries (modulo constraints on syllabic length). Moreover, other experimental techniques, that rely on lists of words rather than on whole sentences, have shown that babies are probably sensitive to word-sized prosodic units before the age of nine months: Several experiments showed that, at this age, they already know the structure of these units in their language. Thus, Jusczyk, Friederici, Wessels, Svenkerud and Jusczyk (1993b) showed that American infants prefer to listen to lists of English words rather than lists of Norwegian words. This is true even when the words are filtered, which shows that infants react to prosodic properties of the words — even though these contain only two syllables. Similarly, Jusczyk *et al.* (1993a) showed that American nine-month-olds prefer to listen to lists of bisyllabic English words that begin with a strong syllable, which is the most frequent stress pattern in English (Cutler and Carter 1987). More direct evidence comes from a recent study by Jusczyk (1995). They showed that seven-and-a-half-month-olds were able to recognize isolated monosyllabic words to which they had been exposed in sentence context. Because this study involved only monosyllabic words, it was premature to

conclude that infants had actually extracted words from the sentences, rather than just isolated syllables. However, recent studies using the same technique indicate that infants are able to extract multisyllabic units from continuous speech passages.

3.2. *Perceptibility of prosodic cues by newborn infants*

The studies presented in the previous section tell us at what age infants gain access to different kinds of prosodic units, but not how they acquired this knowledge. If we are right in assuming that prosodic information is used for acquisition, then very young infants should *at least* be able to perceive and represent this type of information. Sensitivity to prosodic cues is a necessary condition to their use. In order to test whether newborn infants are sensitive to prosodic boundary cues,[11] we tested their ability to discriminate bisyllabic contexts that either contained or did not contain a word boundary (Christophe, Dupoux, Bertoncini and Mehler 1994). We thus compared syllable pairs that were extracted from the middle of a word (for example, *mathé*ma*ticien*), or consisted of the last syllable of a noun and the first of the following adjective (for example, *panora*ma ty*pique*), which corresponds at least to a clitic group boundary, and maybe also to a phonological phrase boundary (as in the sentences used for the above-mentioned adult experiments, still following the terminology of Nespor and Vogel 1986).

Measures of the prosodic characteristics of the stimuli showed that the last vowel of a word was significantly longer than a non-final vowel. There was also a significant lengthening of the first consonant of a word. Moreover, adult French subjects were able to learn to categorize the stimuli better than chance. They were also able to generalize this capacity to new stimuli, which shows that they relied on properties common to each category.

We used the non-nutritive sucking method in order to test three-day-old French infants' sensitivity to cues that differentiate stimuli with and without a prosodic boundary.[12] Sucking rates per minute for experimental and control infants are shown in figure 5.

---

11. It may be the case that newborn infants are not sensitive to prosodic boundary cues, but that this sensitivity appears spontaneously through maturation during the first few months of life. However, research conducted during the last 20 years has shown that even newborns are very well equipped to process speech: They discriminate between sentences from their native language and sentences from a foreign language (Mehler, Jusczyk, Lambertz, *et al.* 1988); they hear the difference between very short CV syllables that differ in only one distinctive feature (Bertoncini, Bijeljac-Babic, Blumstein and Mehler 1987); and they *count* syllables (or maybe vowels, Bijeljac-Babic, Bertoncini and Mehler 1993). These data make it plausible to think that sensitivity to boundary cues may be present from birth.
12. This method consists of presenting infants with one auditory stimulus for each high-amplitude suck. During the habituation phase, infants of the experimental group hear a series of *mati* belonging to one category, and they are switched to stimuli from the other category during the

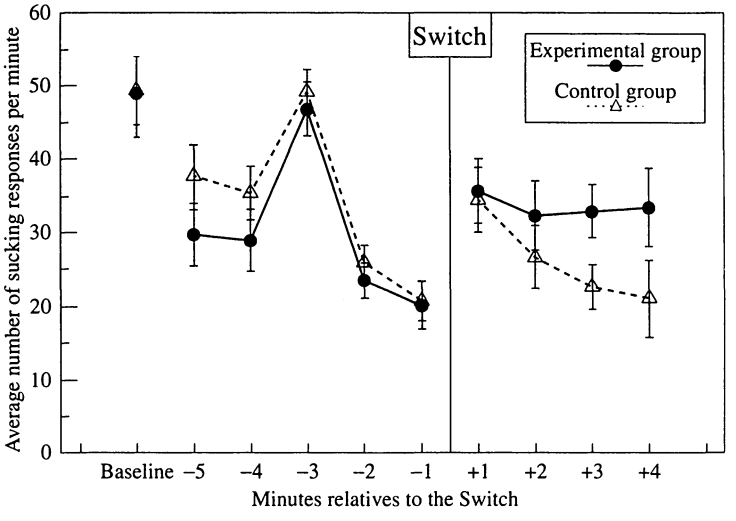

Figure 5. Mean sucking rates per minute for babies in the experimental condition (that change stimuli category) and in the control condition (who go on with stimuli from the same category), *for mati* stimuli with or without a word boundary (from Christophe *et al.* 1994)

One can see that the two groups of infants have similar sucking rates before the moment marked *switch*. After his moment, the experimental group is presented with stimuli from another category, while the control group goes on hearing stimuli from the same category. Infants in the experimental group significantly increase their sucking rates after this point, relative to infants in the control group. This means that three-day-old infants are sensitive to the difference between stimuli with and without a prosodic boundary. This experiment has been replicated with a new phonetic context (for example, *règle*menta*tion* versus *ser*ment ta*cite*) and with a new speaker (the experimental procedure and results are fully reported in Christophe *et al.* 1994). These results allow us to conclude that at least in French, something perceptible happens at prosodic boundaries of the type studied here.

Of course, according to the prosodic segmentation hypothesis, infants should be sensitive to boundary cues in any language of the world, not just in French. If they are to discover which cues mark the prosodic boundaries in their native language, they should be able to perceive all cues that could possibly mark

---

test phase. Control infants, in contrast, are switched to another series of *mati* that are physically different from the ones heard during the habituation phase, but belong to the same category. Thus, any change in the behavior of experimental infants relative to control infants can be due only to the change in category itself (not to the change in actual stimuli). Discrimination has taken place when experimental in infants suck more than control infants during, but not before, the test phase.

boundaries in any language of the world. It is therefore necessary to replicate the preceding study in other languages. In French, there is considerable word-final vowel lengthening due to the word-final accent. But accent is not fixed in every language of the world. We therefore have to check that the above result remains true in a language where accent does not mark boundaries. To do so, we used Spanish, a language with free accent, and we constructed bisyllabic items with the stress always falling on the last syllable (for instance, *latí* extracted from *corre*lati*vamente* or from *Manue*la t*ímida*). Preliminary results show that three-day-old French newborns react to the presence of a boundary in these conditions. An analysis of the prosodic characteristics of the stimuli reveals, just as in French, a significant lengthening of word-initial consonants. Unlike in French, however, there is no significant lengthening of the word-final vowels, although these are significantly higher-pitched than word-medial vowels. Thus, prosodic boundary cues for French and Spanish are similar in some ways but different in others.

To conclude, we have seen that infants acquire some knowledge of the structure of the prosodic units of their maternal language during the first year of life (the series of experiments conducted by Jusczyk and his colleagues). Furthermore, we have shown that newborn infants are already sensitive to prosodic boundary cues in two different languages (French and Spanish). Taken together, these results favor the hypothesis that prosody is used to bootstrap acquisition.

4. Conclusion: A prosody-based model for lexical access and lexical acquisition

In this paper we have raised the question of how the continuous speech stream is segmented into words. We first examined how contemporary psycholinguistic models of lexical access solve the segmentation problem, without the help of an explicit segmentation stage. These models use a mechanism based on multiple activation and competition, that retrieves strings of words that do not overlap. Importantly, these models exploit phonemic information only. Turning to the problem of acquisition, we have seen that such a mechanism cannot be used to acquire a lexicon (a repertoire of word forms), since it crucially rests on knowledge of the language lexicon; therefore, other mechanisms have to be used to this end. Current proposals rest on the use of distributional regularity, phonotactics, and lexical biases (namely the strong initial bias proposed by Anne Cutler for English). Still, none of these mechanisms offer a complete solution to the segmentation-into-words problem, and apart from distributional regularities, they all need at least some boundaries to be bootstrapped. We suggested that prosody would be an invaluable source of information in that it might allow listeners to segment the speech stream into linguistically relevant units.

In the remainder of this conclusion, we come back to the model introduced in figure 2 and speculate on how the unspecified parts of the model may be

filled in. Actually, the majority of results reviewed have given us information about the existence of a prosodic segmentation stage. Thus, we have gathered some evidence that adults exploit prosodic cues in their on-line processing of spoken sentences. As far as babies are concerned, we have shown that they acquire a knowledge of what is a well-formed prosodic unit in their native language during the first year of life, and that newborn infants are already sensitive to prosodic segmentation cues. However, we still know very little about the way in which prosodic segmentation is actually implemented. A possible fully spelled-out model is presented in figure 6, for illustrative purposes: The full boxes and arrows refer to the system as it may work for adults, and the dotted boxes and arrows correspond to mechanisms and routes that are used during acquisition.

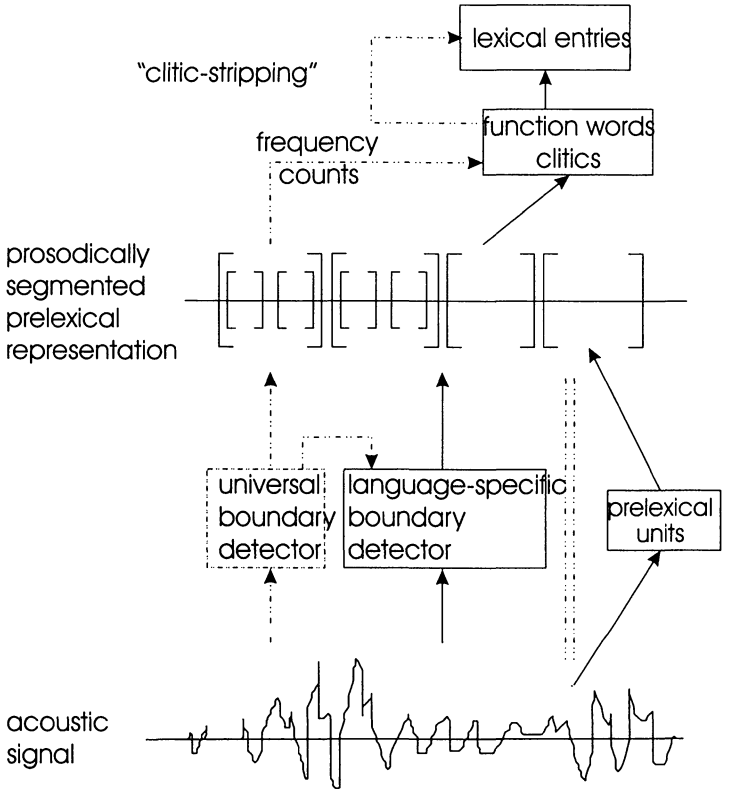

Figure 6. A possible fully-fledged version of a common model for lexical access and lexical acquisition, that rests on the prosodic segmetation hypothesis
Key: Full arrows and boxes refer to the system as it would work in adults, dashed arrows and boxes to the mechanisms and pathways that would be used during the acquisition.

At least two aspects of the model remain to be worked out: The detailed nature of the prosodic units, and how lexical access works within these units. The definition of the prosodic units, in terms of the correspondence with constituents from the prosodic hierarchy has implications for the acquisition of phonology (because prosodic constituents are the domains of phonological rules, having access to them should ease acquisition of the rules). Moreover, because of the relationship between the morpho-syntactic hierarchy and the prosodic hierarchy, knowing prosodic boundaries would give information as to syntactic constituents (for instance, phonological phrase boundaries coincide with left boundaries of X″-constituents — in languages with right recursivity, and right boundaries in languages with left recursivity). This information would be useful to infants for acquiring syntax and to adults for parsing sentences. For instance, Nespor (1995) offered a detailed proposal of how the left–right recursivity parameter might be set through knowledge of the prosodic hierarchy (see also Nespor, Guasti and Christophe in press). It is as yet an open question whether, only one type of prosodic boundary is marked, or whether a hierarchy of units is available. Although we have seen evidence that boundaries corresponding to phonological and intonational phrases (*modulo* performance constraints) may be marked, this does not mean that each perceived boundary can be classified as one or the other. It could be the case for instance that only one basic prosodic level is readily available, although smaller or less reliable prosodic cues may help segmentation within these basic units, along with for example, phonotactics or lexical knowledge.

Also, we will have to address the issue of language specificity in prosodic units. Some prosodic boundary cues have been measured in all the languages that have been studied up to now. Thus, a word-initial consonant lengthening has been reported in all languages studied thus far, and pre-boundary lengthening is also a common phenomenon. More generally, it may be the case that an exaggerated interval between two successive P-centers (for *perceptual-centers*, see Scott 1993, for a recent definition) is a universal marker of a prosodic boundary, although this remains to be demonstrated with a larger number of languages. Some other cues may be language-specific. Thus, we may conjecture that infants might bootstrap their prosodic segmentation strategy by exploiting universal cues: They would discover language-specific cues as correlates of the universal ones. Since we know that babies acquire a relatively precise knowledge of the shape of the prosodic units in their native language during the first year of life, we would expect them to possess a relatively efficient prosodic segmentation procedure, similar to the one for adults, before they reach the end of the first year of life. With regard to the second point, we have to explicit how lexical access and acquisition are accomplished on the basis of the prosodic units (top part of the model in figure 6). Since it is quite plausible that each marked prosodic unit contains several words or morphemes, infants and adult still face the problem of identifying some words without knowing their boundaries. However, we want to stress that prosodic segmentation modifies the

problem quite dramatically — not only quantitatively (because fewer boundaries have to be discovered) but also qualitatively. Indeed, function words tend to appear at the borders of prosodic units. This is certainly true of clitics, which appear most of the time at the beginning or end of clitic groups, with the rare exception of enclitics. Thus, if babies counted the frequency of occurrence of the first and last syllables of prosodic units, they would in all probability find the grammatical morphemes on top of the list. This done, they may *strip* these frequent syllables and construct lexical entries of content words with what is left. The algorithm just described is based on distributional regularity, but it exploits the specific distribution of the very frequent function words. In support of this idea, Gerken and McIntosh (1993) showed that 18-months-old American infants know the function words of English (although they do not pronounce them yet). More crucially, Gerken also showed, using the event-related potential technique, that American infants start recognizing the function words of English at the age of 11 months (Gerken, personal communication). It is thus plausible that function words, and grammatical morphemes in general, may be used by infants in the process of constructing the lexical entries — and in lexical access by adults (a suggestion already made in the literature, see, for example, Garrett 1975).

This particular model is only a sketch and may prove to be wrong in the near future. Only experimental data will decide this. However, we would like to stress two of its features that we feel particularly important. First, it tries to explain acquisition by infants and processing by adults within the same framework. As we mentioned in the introduction, we think that the highly efficient, language-specific, processing system used by adults when perceiving speech derived from the universal architecture used by infants in the process of acquisition. This position puts constraints on both parts of the model. Thus, for each proposed mechanism, an account of how the necessary knowledge can be acquired has to be offered. In addition, the model can be constrained by experimental data gathered from infants as well as adults. Second, the model tries to incorporate prosodic information, which is usually left out from psycholinguistic models, probably for want of an adequate framework. We feel that prosody is an invaluable source of information (bearing on many aspects of language, not just lexical access) that could considerably enrich existing models. We think that prosodic phonology offers a much-needed framework from which we can start studying the role of prosody in speech perception and production. Whatever the fate of this particular model, we hope that the underlying ideas will prove successful.

Received 14 June 1994
Revised 28 July 1995

(Christophe)
LSCP EHESS-CNRS, Paris, and
MRC Cognitive Development Unit, London
(Dupoux)
LSCP EHESS-CNRS, Paris, and
France Telecom, Paris

References

Altmann, G. (1990). *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives.* Cambridge, MA: MIT Press.

Bertoncini, J., R. Bijeljac-Babic, S. Blumstein and J. Mehler (1987). Discrimination of very short CV syllables by neonates. *Journal of the Acoustical Society of America* 82: 31–37.

Bijeljac-Babic, R., J. Bertoncini and J. Mehler (1993). How do four-day-old infants categorize multisyllabic utterances? *Developmental Psychology* 29: 711–721.

Bolinger, D. (1978). Intonation across languages. In *Universals of Human Language,* Joseph Greenberg (ed.), 471–524. Stanford: Stanford University Press.

Brent, M., T. Cartwright and A. Gafos (in press). Distributional regularity and phonotactics constraints are useful for segmentation. *Cognition.*

Cairns, P, R. Shillcock, N. Chater and J. Levy (in press). Bootstrapping word boundaries: A bottom–up corpus-based approach to speech segmentation. *Cognitive Psychology.*

Christophe, A. (1993). Rôle de la prosodie dans la segmentation en mots. Unpublished Ph.D. dissertation, EHESS, Paris, France.

Christophe, A., E. Dupoux, J. Bertoncini, J. Mehler (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America,* 95: 1570–1580.

Christophe, A., C. Pallier, E. Dupoux and J. Mehler. Hearing words inside words? A phoneme-monitoring study of lexical activation. Unpublished manuscript. Submitted.

Cole, R. and J. Jakimik (1980). A model of speech perception. In *Perception and Production of Fluent Speech,* R. Cole (ed.), 133–163. Hillsdale, NJ: Erlbaum.

Cruttenden, A. (1986). *Intonation.* Cambridge: Cambridge University Press.

Cutler, Ann (1990). Exploiting prosodic probabilities in speech segmentation. In *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives,* G. Altmann (ed.), 105–121. Cambridge, MA: MIT Press.

Cutler, Ann (1996). Prosody and the word boundary problem. In *From Signal to Syntax,* J. Morgan and Katherine Demuth (eds.), 87–99. Hillsdale, NJ: Lawrence Erlbaum.

Cutler, Ann and D. Carter (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2: 133–142.

Cutler, Ann and D. Norris (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14: 113–121.

Cutler, Ann and S. Butterfield (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31: 218–236.

Cutler, Ann and J. Mehler (1993). The periodicity bias. *Journal of Phonetics* 21: 103–108.

Delais-Roussarie, E. (1995). Pour une approche parallèle de la structure prosodique: Etude de l'organisation prosodique et rythmique de la phrase française. Ph.D. dissertation, Univerity of Toulouse-Le Mirail, France.

Dirksen, A. (1992). Accenting and deaccenting: A declarative approach. In *Proceedings of the 15th International Conference on Computational Linguistics, COLING '92, (Nantes, France: ICCL)* 3: 865–869.

Dupoux, E., A. Christophe, J. Mehler. Lexical effects in phoneme monitoring: Time course versus attentional accounts. Unpublished manuscript.

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

Foss, D. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior* 8: 457–462.

Frauenfelder, Uli (1991). Lexical alignment and activation in spoken word recognition. In *Music, Language, Speech and Brain,* J. Sundberg, L. Nord and R. Carlson (eds.), 294–303. London: Macmillan.

Friederici, A. and J. Wessels (1993). Phonotactic knowledge of word boundaries and its use in infant speech-perception. *Perception and Psychophysics* 54: 287–295.

Garrett, M. (1975). The analysis of sentence production. In *The Psychology of Learning and Motivation: Advances in Research and Theory,* Vol.9, G. Bower (ed.), 133–177. New York: Academic Press.

Gee, J. and F. Grosjean (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15: 411–458.

Gerken, L., P. Jusczyk and D. Mandel (1994). When prosody fails to cue syntactic structure: 9–Month-olds' sensitivity to phonological versus syntactic phrases. *Cognition* 51: 237–265.

Gerken, L. and B. McIntosh (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology* 29: 448–457.

Goodsitt, J., J. Morgan and P. Kuhl (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language* 20: 229–252.

Gow, D. and P. Gordon (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance* 21: 344–359.

Grosjean, F., L. Grosjean, H. Lane (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology* 11: 58–81.

Gussenhoven, Carlos and A. Rietveld (1992). Intonation contours, prosodic structure and pre-boundary lengthening. *Journal of Phonetics* 20: 282–303.

Harrington, J. and A. Johnstone (1987). The effects of equivalence classes on parsing phonemes into words in continuous speech recognition. *Computer Speech and Language* 22: 273–288.

Hayes, Bruce and H. Clark (1970). Experiments on the segmentation of an artificial speech analogue In *Cognition and the Development of Language,* B. Hayes (ed.), 221–234. New York: Wiley.

Hirsh-Pasek, K., D. K. Nelson, P. Jusczyk, K. Cassidy, B. Druss and L. Kennedy (1987). Clauses are perceptual units for young infants. *Cognition* 26: 269–286.

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

Jusczyk, P. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* 29: 1–23.

Jusczyk, P., A. Cutler and N. Redanz (1993a). Infants' preference for the predominant stress patterns of English words. *Child Development* 64: 675–687.

Jusczyk, P., A. Friederici, J. Wessels, V. Svenkerud and A. Jusczyk (1993b). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language* 32: 402–420.

Jusczyk, P., D. Kemler-Nelson ,K. Hirsh-Pasek, L. Kennedy, A. Woodward and J. Piwoz (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology* 24: 252–293.

Jusczyk, P., D. Kemler-Nelson, K. Hirsh-Pasek and T. Schomberg (1995). Perception of acoustic correlates to clausal units in a foreign language by American infants.

Jusczyk, P. and C. Krumhansl (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance* 19: 627–640.

Lahiri, Aditi and W. Marslen-Wilson (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38: 254–294.

Lehiste, Ilse (1965). Juncture. *Proceedings of the Fifth International Congress of Phonetic Sciences* 1964: 172–200.

—— (1966). Consonant quantity and phonological units in Estonian. [=Indiana University Publications of the Uralic and Altaic Series, 65.] The Hague: Mouton.

Luce, P. (1986). A computational analysis of optimal discrimination points in auditory word recognition. *Perception and Psychophysics* 39: 155–159..

Marslen-Wilson, W. and P. Warren (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review* 101: 653–675.

McClelland, J. and J. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology* 18: 1–86.

McQueen, J. and A. Cutler (1992). Words within words: Lexical statistics and lexical access. In *Proceedings of the International Conference on Spoken Language Processing,* 1: 221–224. University of Alberta, Banff, Alberta, Canada.

McQueen, J., D. Norris and Ann Cutler (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 621–638.

Mehler, J., J. Bertoncini, E. Dupoux and C. Pallier (1994). The role of suprasegmentals in speech perception and acquisition. *Dokkyo International Review* 7: 343–377.

Mehler, J., E. Dupoux, T. Nazzi and G. Dehaene-Lambertz (1996). Coping with linguistic diversity: The infant's viewpoint In *From Signal to Syntax,* J. Morgan and K. Demuth (eds.), 101–116. Hillsdale, NJ: Erlbaum.

Mehler, J., E. Dupoux, J. Segui (1990). Constraining models of lexical access: The onset of word recognition. In *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspective,* G. Altmann (ed.), 263–280. Cambridge, MA: MIT Press.

Mehler, J., P. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini and C. Amiel-Tison (1988). A precursor of language acquisition in young infants. *Cognition* 29: 143–178.

Monnin, P. and F. Grosjean (1993). Les structures de performance en français: Caractérisation et prédiction. *l'Année Psychologique* 93: 9–30.

Morgan, J. and K. Demuth (1996). *From Signal to Syntax.* Hillsdale, NJ: Lawrence Erlbaum.

Nakatani, L., K. O'Connor and C. Aston (1981). Prosodic aspects of American English speech rhythm. *Phonetica* 38: 84–106.

Nakatani, L. and J. Schaffer (1978). Hearing words without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America* 63: 234–245.

Nespor, M. (1995). Setting syntactic parameters at a prelexical stage. *Proceedings of the XXV ABRALIN Conference Salvador deBahia.*

Nespor, M., M. Guasti and A. Christophe. Selecting word order: The rhythmic activation principle. In *Interfaces in Phonology*, M. Bierwisch and U. Kleinhenz (eds.). Berlin: Akademie Verlag. In press.

Nespor, M. and I. Vogel (1986). *Prosodic Phonology.* Dordrecht: Foris.

Newsome, M. and P. Jusczyk (1995). Do infants use stress as a cue in segmenting fluent speech? In *Proceedings of the 19th Boston University Conference on Language Development,* C. MacLaughlin and S. McEwen (eds.) 2: 415–426 Boston, MA: Cascadilla Press.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52: 189–234.

Norris, D., J. McQueen and A. Cutler. Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21: 1209–1228.

Pijper, J. de, and A. Sanderman (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America* 96: 2037–2047.

Pisoni, D. and P. Luce (1986). Speech perception: Research, theory, and the principal issues. *Pattern Recognition by Human and Machines: Speech Perception* 1: 1–157.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics* 20: 331–350.

Rietveld, A. (1980). Word boundaries in the French language. *Language and Speech* 23: 289–296.

Rubin, P., M. Turvey and P. V. Gelder (1976). Initial phonemes are detected faster in spoken words than in nonwords. *Perception and Psychophysics* 19: 394–398.

Scott, S. (1993). Perceptual centers in speech and acoustic analysis. Unpublished Ph.D. dissertation, University College London, London.

Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge, MA: MIT Press.

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In *Cognitive Models of Speech Processing: Psycholinguistic and computational Perspectives,* G. Altmann (ed.), 24–49. Cambridge, MA: MIT Press.

Swinney, D., W. Onifer, P. Prather and M. Hirshkowitz (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory and Cognition* 7: 165–195.

Swinney, D. (1981). Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In *The Cognitive Representation of Speech,* T. Myers, J. Laver and J. Anderson (eds.), 201–209. Amsterdam: North-Holland.

Tabossi, P. (1993). Connections, competitions, and cohorts: Comments on the chapters by Marslen-Wilson; Norris; and Bard and Shillcock. In *Cognitive Models of Speech Processing: The Second Sperlonga Meeting,* G. Altmann, and R. Shillcock (eds.), 277–294. Hillsdale, NJ.: Lawrence Erlbaum.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America* 61: 846–858.

Vaissière, J. (1983). Language-independent prosodic features. In *Prosody: Models and Measurements.* A. Cutler and D.R. Ladd (eds.), 53–66. Berlin: Springer Verlag.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32: 25–64.