Cognition, 25 (1987) 135–155

Prosodic structure and spoken word recognition

FRANÇOIS GROSJEAN*
Northeastern University

JAMES PAUL GEE
Boston University

Abstract

The aim of this paper is to call attention to the role played by prosodic structure in continuous word recognition. First we argue that the written language notion of the word has had too much impact on models of spoken word recognition. Next we discuss various characteristics of prosodic structure that bear on processing issues. Then we present a view of continuous word recognition which takes into account the alternating pattern of weak and strong syllables in the speech stream. A lexical search is conducted with the stressed syllables while the weak syllables are identified through a pattern-recognition-like analysis and the use of phonotactic and morphonemic rules. We end by discussing the content word vs. function word access controversy in the light of our view.

1. Introduction

The aim of this paper is to present a view of lexical access that takes into account the prosodic structure of spoken language. We first argue that current spoken word recognition models are overly influenced by the written language notion of the word. We then discuss aspects of prosodic structure, and stress the importance that should be given to them in models of language pro-

*The authors are co-equal and only the laws of alternation are responsible for the order of the names. Much of the thinking behind this paper was done at the Max Planck Institute, Nijmegen, The Netherlands, while the authors were visiting research fellows there in July 1983. They are grateful to all those in the Institute who made their stay there so fruitful. This work was also supported in part by grants from the National Science Foundation (BNS-8404565), The National Institute of Health (RR 07143) and the Northeastern University RSDF funds. The authors are most grateful to Bill Cooper, Jean-Yves Dommergues, Uli Frauenfelder, Harlan Lane, Jacques Mehler, Joanne Miller and Lorraine Tyler for their helpful comments on a first draft of the paper. Very special thanks also go to three anonymous reviewers who, through their comments, have helped improve this paper. Requests for reprints should be sent to François Grosjean, Department of Psychology, Northeastern University, Boston, MA 02115, U.S.A.

cessing. Next we present a view of lexical access that parses the phonetic and syllabic representation of the speech stream into words. This is done through two types of analysis. On the one hand, the processing system searches for stressed syllables and uses these to initiate a lexical search. On the other hand, the system identifies the weak syllables on either side of the stressed syllable by subjecting them to a pattern-recognition-like analysis and by using the phonotactic and morphophonemic rules of the language. These two types of analysis interact with one another and with other sources of information to segment the speech stream into a string of words. We end by suggesting that the existing distinction between function words and content words in the domain of lexical access can be better understood in terms of the two types of analysis we propose.

2. The problem

Considerable advances have been made in the understanding of the processes that underlie spoken (as opposed to written) word recognition. We know, for example, that certain properties of words affect their recognition: their frequency of use (Howes, 1957; Rubenstein & Pollack, 1963); their length (Grosjean, 1980; Mehler, Segui & Carey, 1968); their phonotactic configuration (Jakimik, 1979) and their uniqueness point (Marslen-Wilson, 1984). We also know that when words are presented in context, their lexical properties interact with various sources of knowledge (linguistic rules, knowledge of the world, discourse, etc.) to speed up or slow down the recognition process (Cole & Jakimik, 1978; Grosjean, 1980; Marslen-Wilson & Welsh, 1978; Morton & Long, 1976; Tyler & Wessels, 1983). The exact nature of the "interaction" between the properties of the words and these sources of knowlledge remains to be described adequately, and the controversy concerning the moment at which "top-down" information enters the lexical access process has yet to be resolved (Forster, 1976; Marslen-Wilson & Welsh, 1978; Morton, 1969; Swinney, 1982). One conclusion that emerges from this research is that recognizing a word may not be a simple mapping between its acoustic-phonetic properties and its entry in the mental lexicon (although, see Klatt, 1979). Instead, it may well be a rather complex process that involves various narrowing-in and monitoring stages, correcting strategies, post-access decision stages, and even look-ahead and look-back operations (Grosjean, 1980; Marslen-Wilson & Welsh, 1978; Nooteboom, 1981; Swinney, 1982).

Despite these advances in the field of word recognition, there is one problem that may be undermining current research—it is the over importance given to the written language notion of the word. Many researchers have assumed that the written word has an acoustic-phonetic correlate in the speech stream and that it is this "unit" that is involved in the recognition process. Whether there is direct access from the spectra (as in Klatt, 1979) or intermediary stages of representation that involve linguistic units such as the phoneme or the syllable (as in most other models), the assumption has been that the domain over which processing takes place is the spoken analog of the written dictionary word (the WD word). We agree that the final outcome of spoken word recognition is the stored lexical item in the internal lexicon (an item that shares many characteristics with the WD word)[1] but we question whether, in the earlier stages of lexical access, the WD word plays the important role many have assumed: either as a "unit of processing" or as the domain over which lexical access takes place.

Several reasons can explain the importance given to the WD word in the lexical access of spoken words. First, spoken word recognition research has always lagged behind the research on written word recognition and has therefore borrowed from it its unit of study—the WD word. Because written language is ever present in our everyday life, it is all too easy to think of spoken language as a concatenation of individual words, even though the actual acoustic-phonetic stream of spoken language does not reflect this. Second, research has usually investigated the recognition process of single words, presented in the speech stream or in their canonical form, and not the operations involved in continuous lexical access. The consequence of this has been to strengthen the importance given to the individual WD word during the recognition process. Third, the experimental tasks used in spoken word recognition are too often biased towards single WD-like words; thus, word monitoring and lexical decision, among other tasks, force the listener to focus on single words; this, in turn, encourages us to think that the human processing system may in fact use the WD word as a unit or domain of processing. Finally, many of the words that are used in experiments are content words (as opposed to function words) and have few, if any, inflections. One consequence of this is that they are often recognized before their acoustic-phonetic offset, and this result reinforces the view that the spoken word recognition

[1]Even this question is more complex than it might seem at first. In the case of words put together by productive morphology, the final outcome of spoken word recognition may not be the word itself but a set of morphemes. These morphemes have to be related to each other in some fashion—often hierarchically with one morpheme being taken as the "head" around which the rest of the word is organized—before the whole word is "recognized". Particularly interesting problems arise in the case of polysynthetic languages, where words are made up of large numbers of morphemes put together in quite productive ways and where sentences contain few words (Comrie, 1981). For the hierarchical structure of words in English, see Selkirk (1982) and Williams (1981).

system is WD word driven, from the segmentation of the speech stream to the final access of the lexicon.

The importance given to WD words in lexical access research has led to a number of critical assumptions about spoken word recognition. One that is assumed by many researchers is that words are recognized sequentially, left to right, one word at a time. Cole and Jakimik (1979, pp. 133–134) state this explicitly when they write:

> Speech is processed sequentially, word by word ... the words in an utterance are recognized one after another ... listeners know where words begin and end by recognizing them in order.

Likewise, Marslen-Wilson and Welsh (1978), Morton (1969) and Forster (1976), in their different models of word recognition, assume word by word processing. Thus, in the cohort model, the first segment of a WD word plays the critical role of activating a cohort of candidates; in the logogen model, each WD word has a logogen, and in the bin model, one accesses the master lexicon by first finding an entry for the WD word in question. There is ample evidence that many words are indeed recognized sequentially (Grosjean, 1980; Marslen-Wilson & Tyler, 1980; Tyler and Wessels, 1983), but as we will argue below, this may not be the case for ALL words, and word by word recognition models will have to account for this. (Note that more recent models, like that of McClelland & Elman, 1986, do not make the sequentiality assumption.)

Other assumptions fall out of this WD word by WD word approach. One is that the beginning of the word is critical in its recognition (but what happens if the system does not know it is dealing with the beginning of a word? See Marslen-Wilson & Welsh, 1978; Cole & Jakimik, 1979, for a discussion of this assumption), and another assumption is that words are recognized at that point in their acoustic-phonetic configuration where they distinguish themselves from every other word (again proposed by Marslen-Wilson & Welsh, and Cole & Jakimik). The above assumptions may be correct for many content words (especially if they occur in the appropriate context) but may not always apply to certain monosyllabic and low frequency words, unstressed function words, words stressed on the second syllable, and words with prefixes and suffixes.

One need only examine the spectrogram or waveform of an utterance to be reminded of the continuous nature of the speech stream and of the ever present segmentation problem faced by the speech perception and word recognition systems. There is, in fact, evidence in the literature which shows that WD words are not all recognized one at a time in a sequential manner. In a set of classic studies, Pollack and Pickett (1963, 1964) found, for example, that words extracted from the speech stream were not easily recognized by themselves. Only 55% of the words were identified correctly when presented in isolation, and this percentage only reached 70–80% accuracy when the words were presented in two or three word samples. Grosjean (1980) found similar results in a gating study when he extracted WD words from the speech stream and presented them to subjects from left to right in increments of increasing duration: only 50% of the one syllable low frequency words (and 75% of the high frequency words) were guessed correctly by 5 or more of the 8 subjects used. (See Cotton & Grosjean, 1984, and Tyler & Wessels, 1985, for evidence showing that the gating paradigm shares many of the characteristics of other on-line tasks.) These studies appear to show, therefore, that not all WD words are recognized before their acoustic offset, a claim that is made (explicitly or implicitly) by many current models of word recognition.

To show that these results were not due to the fact that the words were extracted from their natural context, Grosjean (1985) constructed sentences of the type:

(1) I saw the bun in the store
(2) I saw the plum on the tree

and gated from the beginning of the object noun ("bun", "plum") all the way to the end of the sentence. The subjects were asked to listen to "I saw the", guess the word being presented after it, indicate how confident they were of their guess, and finish off the sentence. Whereas all two- and three-syllable words were isolated (guessed correctly) before their acoustic offset, only 45% of the one-syllable words were identified correctly. The remainder were isolated during the following preposition or even during the article after that. An examination of the point at which subjects felt perfectly confident of their guess confirmed the isolation results. Whereas 80% of the two- and three-syllable words were given perfect confidence ratings before their acoustic offset, only 5% of the one-syllable words managed this; 50% were given perfect ratings during the next word, 20% had to wait for the following article, and 25% were only confirmed during the perception of the following noun—some 500 ms after the onset of the stimulus object noun! An examination of the erroneous guesses proposed by subjects before they isolated the appropriate word confirmed the isolation times and the confidence ratings. When presented with the acoustic information of the next WD word (the preposition), subjects often thought they were dealing with the second syllable of the stimulus word (e.g., "bunny", "bonnet" for "bun in"; "boring" for "boar in"; "plumber" for "plum on", etc.). These erroneous guesses remind one of the slips of the ear reported in the literature. For example, Bond and Garnes

(1980) state that some 70% of the multiple word slips they found concerned word boundary shifts (an ice bucket → a nice bucket), word boundary deletions (ten year party → tenure party), and word boundary additions (descriptive linguistics → the script of linguistics).

The WD word (as a linguistic unit in its own right, as an ensemble of smaller linguistic units, or as an acoustic-phonetic segment) may not after all be the ideal domain over which lexical processing takes place. WD words are rarely delineated clearly in the acoustic-phonetic stream and many may not always be processed by the listener before their acoustic offset. Some form of the WD word must be stored in the lexicon (but see footnote 1) and as such is the final product of the word recognition process (as listeners, we need to know which words have been said), but we question whether the early stages of lexical access involve strictly left to right, WD word by WD word processes. The problem is of course compounded when we reflect on how function words, which are often short, unstressed and reduced phoneticcally, are recognized on-line. It is doubtful whether strings of unstressed function words such as "could have been", "to the", "I'd have" are in fact recognized WD word by WD word, in a left to right sequential manner. Similar doubts can be expressed about content words that have undergone "deaccenting" in discourse context (Ladd, 1980).

Before presenting a view of word recognition that puts less stress on the left to right, WD word by WD word, process, it is important to discuss certain characteristics of the prosodic (phonological) structure of spoken language that may well play an important role in lexical access. This we do in the section below.

3. Prosodic structure and phonological words

Recent prosodic (phonological) theories of language propose that the utterance has a suprasegmental, hierarchical organization defined in terms of a metrical tree with binary branching and strong/weak prominence relations. These prominence relations are defined for all the elements that make up the units of the prosodic structure, from the syllable through the word and phrase to the utterance as a whole (Liberman, 1975; Liberman & Prince, 1977; Selkirk, 1978, 1980a,b).

According to one view of prosodic structure (Culicover & Rochemont, 1983; Gee & Grosjean, 1983; Nespor & Vogel, 1982, 1983; Selkirk, 1978, 1980a,b), at the lower level of the hierarchy, weak and strong syllables group together into feet. Feet in turn bundle into phonological words which may, or may not, correspond to traditional WD words. Thus, in the sentence:

(3) John//has been avidly/ reading//the latest/news// from home

we note that some phonological words (marked off by one or two slashes) correspond to WD words (John, reading, news) but that others do not and instead are made up of two or three WD words (has been avidly, the latest, from home). In the latter case, unstressed function words have been appended to stressed content words.[2]

Phonological words, to which we will return below, bundle into phonological phrases which comprise all the material up to and including the head of a syntactic phrase. (Phonological phrases are marked off with two slashes in sentence 3.)[3] Of interest here is that phonological phrases are not always constituents in syntactic structure. Thus, in the above example, the phonological phrase boundaries after "reading" and "news" occur in the middle of syntactic phrases, and the phonological phrases "has been avidly reading" and "the latest news" are not syntactic constituents. Phonological phrases themselves bundle into intonational phrases. We should note that there is a great deal of freedom as to what can be an intonational phrase as this is contingent in part on the information that is put into focus in the discourse (Gee & Grosjean, 1983; Selkirk, 1984).

In a recent study, Gee and Grosjean (1983) showed that the segmentation that takes place during oral reading and that is reflected by silent pauses, can best be accounted for by prosodic structure: phonological words are separated from one another by few (if any) pauses which in turn are very short;

[2]The entity we are calling a "phonological word" is not given a uniform name, description, or treatment in the literature. In Selkirk (1978), a phonological rule, the "Monosyllable Rule", operating with the phonological phrase as its domain, causes monosyllabic words which are weak (unstressed) and which correspond to non-lexical items in the syntax (function words), to lose their word status at the prosodic level. In this case, they become attached to an adjacent strong syllable (the stressed syllable of some content word). They are thus turned into simple weak syllables, and undergo the phonological rules that other weak syllables undergo (e.g., the weak syllables of multisyllabic words). For a detailed characterization of phonological words, also see Kean (1979, 1980) and Garrett and Kean(1980).

[3]The principles determining the boundaries of phonological phrases are discussed in detail in Culicover and Rochemont (1983), Gee and Grosjean (1983), Nespor and Vogel (1979, 1983), and Selkirk (1979, 1980a,b). Phonological phrases represent phonological (in fact, prosodic) readjustments of syntactic structure. In syntactic theory, any phrase is defined in terms of its head (a noun phrase is headed by a noun, a verb phrase is headed by a verb, a prepositional phrase is headed by a preposition, and so forth). A phonological phrase is made up of all the material up to and including the head of a syntactic phrase. Material following the head normally constitutes a separate phonological phrase. There are several provisos however. First, at the level at which phonological phrases are assigned in the grammar, prepositions no longer count as heads of their phrases, since they are functors and have been destressed (or assigned no stress); thus the noun complement following the preposition serves as the final boundary of the phonological phrase. Second, given the way stressless items attach to stressed ones in English, unstressed object pronouns attach to the verb and become part of a phonological phrase with it (e.g., "him" is part of the second phrase in "The girl/ has hit him"). See Gee and Grosjean (1983) for suggestions about the role such details play in the production system.

phonological phrases are separated from one another by more and longer pauses, and intonational phrases are marked off by more numerous and longer pauses. Syntactic structure, with its traditional units (the WD word, the phrase, the clause) could not account for the pausing data as well. The hierarchical "performance" trees that were constructed from the pause data were often quite different from the syntactic trees but were practically identical to the prosodic trees.

Recent developments of prosodic theory have important implications for psycholinguistics, and models of language production, perception and comprehension will need to pay more attention to the prosodic structure of spoken language utterances. In the domain of lexical access, models will need to take into account prosodic units such as the syllable, the foot and the phonological word (Kean, 1979, 1980, for example). The unit that is of primary interest to us at this time is the phonological word. Although we will *not* argue that it is the unit of lexical access (we only have evidence for it as a unit of production), we will present certain of its characteristics that are important to understand our view of spoken word recognition. The phonological word is a tightly bound phonological unit that is made up of one stressed syllable and a number of weak (unstressed) syllables that are phonologically linked to it. The unit is the domain of various phonological rules, many of which tend to assimilate and blend sounds together (Zwicky, 1977, 1982). The weak syllables in the phonological word may be the unstressed syllables of a content word (e.g., "poral" in "temporal"), affixes attached to a stem ("un" and "ful" in "unhelpful"), clitics attached to a content word ("j' " in the French "j'peux"), reduced functors lexically attached to a content word ("have to" = "hafta", "out of" = "outta"), or function words phonologically linked to content words ("a" in "a dog", "him" in "hit him", etc).

What becomes clear from a study of phonological word is that speech is made up of alternations of weak and strong syllables where the weak syllables are always linked to strong syllables in rule governed ways. Thus certain syllables in the speech stream are more *salient* than others. In English, this saliency is marked by a complex interaction of pitch, duration, and amplitude, for which the term "stress" is often used (Bolinger, 1958a, 1965, 1981; Chomsky & Halle, 1968; Halle & Keyser, 1971). In other languages, saliency is physically realized in different ways (e.g., pitch accent, tone, characteristics of syllable structure, etc.). Salient syllables are realized in words that tend to carry the least redundant and most contentful meaning in a sentence (Bolinger, 1958b, 1961, 1965, 1972; Ladd, 1980). It is interesting to note that in historical change, words that carry less informational saliency tend to lose some of their phonological substance, eventually becoming function words or affixes to other words (Givon, 1975). Such historical changes give rise to a greater degree of alternation between more salient and less salient syllables, both in terms of information and in terms of physical realization. We note also that in early child language, and in the early stages of pidgin languages, nearly every word tends to carry stress. As these language forms develop, however, redundant or less salient information tends to get de-stressed; thus, for example, children in normal first language acquisition eventually add more function words, contract non-contracted forms, and de-stress redundant or "given" content words (Bickerton, 1981; Givon, 1979; Slobin, 1977, 1982). In every instance of developed language, then, we find alternations between elements that carry informational saliency and those that do not, and this alternation is often reflected in the speech stream by some form of prosodic marking of weak and strong syllables. It is our belief that these design features of human language should be integrated into models of language processing. In the next section, we propose a view of lexical access that attempts to do just that.[4]

4. A different view of continuous lexical access

A few points need to be made before presenting our view of lexical access. The first is that it is still very general and thus many aspects of it remain unspecified. The aim behind our presentation is not to argue for the replacement of better established models of word recognition with our own view, but rather to invite researchers to consider a proposal for lexical access that takes into account both the prosodic structure of language and the fact that word recognition may not be a strictly left to right, WD word by WD word process. The second point is that little direct and non-controversial evidence for our view is yet available. We will nevertheless present some indirect empirical data to support it. A third point is that several researchers, among them Cutler (1976) and Bradley (1980), have proposed very similar views of lexical access (see below); this makes our view much less "extreme" than some might think at first.

The approach we propose is basically one that parses the string of weak and strong syllables in the speech stream and that makes contact with the WD-like words in the lexicon. The analyses we will describe work on the output of the speech perception system which comes in the form of a phonetic

[4]The particular way in which we will relate lexical access to weak-strong prosodic relations does not give any grounds for choosing between metrical theory (Hayes, 1980, 1982; Kiparsky, 1979; Liberman & Prince, 1977; McCarthy, 1979; Safir, 1979; Selkirk, 1978, 1980b) and grid theory (Prince, 1983; Selkirk, 1984). For convenience we talk throughout in terms of metrical theory.

string marked in terms of weak and strong syllables. Thus unlike Klatt's (1979) model which does lexical access directly from the spectra, we propose an intermediary representation between the acoustic string and the stored lexical item: it is a string of phonetic segments grouped into syllables marked as weak or strong.[5] We have little to say at this time about the mapping that takes place between the acoustic stream and this intermediate level of representation. What we are concerned with here is the process that parses this representation in such a way as to ultimately make contact with the lexicon. Our view is as follows. On the one hand, stressed syllables (and only they) are used to initiate a lexical search. On the other hand, and concurrently, the weak syllables located on either side of the stressed syllable (functors, affixes, weak syllables of content words) are identified by means of a pattern-recognition-like analysis and with the help of the listener's knowledge of phonotactic and morphophonemic rules.

The actual work that is done with the stressed syllable during the lexical search is as yet unclear: there may be cohort activation, active search of a word, word detector triggering, etc. One possibility that comes to mind is that a series of cohorts are activated where the stressed syllable in question is the first syllable of a subset of candidates, the second syllable of another series of candidates, the third syllable of yet another subset, etc. The information that comes back from the lexical search will be used in two ways: it will help the system recognize the word which contains the stressed syllable and it will play a role in identifying the weak syllables (functors, affixes, etc.) on either side of the stressed syllable.

As a lexical search is taking place with the stressed syllable, the system is trying to identify the weak syllables around it. The identification process is done by bringing to bear the phonotactic and morphophonemic rules of the language as well as by doing a pattern-recognition-like analysis of weak (and weak-strong) syllable groups. We hypothesize here that the listener will be helped by his/her knowledge of well learned syllable patterns such as sequences of function words or beginnings of phonological words (for example, /ðə-----/, /ən----/, /təðə-----/, /ɑɪdəv-------/). The system will also use all the information that comes back from the lexical search that is being conducted with the stressed syllable.

We should note at this point that the analysis of weak syllables may at

[5]We should point out that this intermediate representation need not always be as rich as this. In fact, a representation in terms of sonority of segments, syllable structure, and weak-strong relations would already be quite rich. We should further note that as our view becomes more refined, it may well use the hierarchical structure imposed on weak-strong relations by current metrical theories to much greater purpose than we do here.

times be able to identify certain functors[6] before getting feedback from the lexical search on the candidates containing the stressed syllable and thus indirectly on the unstressed syllables surrounding it. This is the case, for example, when the rules of the language mandate a particular function word or when the functors to the left and right of the stressed syllable cannot make up a possible WD word with it (note, for example, that /ð/ as in "the" or "this" only occurs in functors). This may also be the case with strings of functors that have rather precise and well-known phonetic patterns such as "could have been" (/kudəbɪn/), "I would have" (/aɪdəv/); these may be identified as they are heard. Thus, there is clearly a top-down aspect to this weak syllable analysis. Knowledge of the phonotactic rules of the language gives rise to decisions about which weak syllables may go with stressed syllables and which may not. Of course, this process can sometimes misinform the lexical search. For example, in the "bun in" → "bunny" example given above, the process sees /bʌnɪn/ as a possible phonotactic combination in English and tells the lexical search (working with /bʌn/) that it is a possible candidate (along with /bʌni/). It is only feedback from the lexical search that tells the system that there is no word that corresponds to /bʌnɪn/, and therefore that two words are involved: "bun" and "in".[7]

The constant interaction between these two types of analysis—the analysis conducted on the weak syllables and the lexical search taking place with the stressed syllable—and the ever present information of other sources of information (the listener's knowledge of the world and of the immediate situation,

[6]Fries (1952) contains one of the classic discussions of function words and content words. Fries points out that function words (he identifies 154 in English) must be recognized in a different way than content words, due to the role they play in the sentence. What at the level of syntax amounts to a difference of function translates at the level of prosodic structure into a difference of stress (or saliency). In both cases, we may have something more akin to a continuum than a categorical distinction. See the discussion in Section 5 below.

[7]It is important to realize that even with partial information about phonological segments, a system utilizing its linguistic knowledge of phonological structure (at the segment, syllable, foot, and phonological levels) can make good predictions about possible word boundaries. Lynn Frazier (1987, this issue), utilizing work by Taft (forthcoming) and Selkirk (1980a), demonstrates how information about the sonority of segments, syllable structure, and phonological word structure predicts, for example, that a string of syllables consisting of "strong-weak-strong-strong" will be initially parsed as two words, the first word having the structure "strong-weak-strong" and the second being a strong monosyllable. In addition, she points out that for such a system to work, it need only have partial information about segments: syllabic parsing (in the sense of parsing the string into syllables) "could proceed before the identity of the segment has been established assuming that the sonority relations between that segment and adjacent segments may be determined in the absence of a complete feature specification (e.g., if the place of articulation of a segment was still unspecified)." It is interesting to note also that sonority is just a weak-strong relation defined internal to a syllable, as stress is a weak-strong relation defined above the level of the syllable. Our hypothesis then is that the process analyzing weak syllables (or groups of weak and strong syllables) is computing phonological structure in order to gain information about boundaries and to aid the lexical search being conducted with the strong syllable.

the rules of the language, the discourse, etc.) will allow the appropriate segmentation of the speech stream into a string of WD-like words.

Two points need to be made at this time. First, whereas words which contain a stressed syllable are recognized as a consequence of a complex lexical search (involving the lexicon at all stages of the search), unstressed words (and especially unstressed functors) are identified in a more direct manner. It is only when these words have been (partly?) identified that direct contact is made with the lexicon in order to obtain the information needed about them (meaning, syntactic class, subcategorization information, etc.). Functors are probably listed in the lexicon by themselves as well as in the general lexicon (see Bradley, 1978); depending on their saliency, they will be "looked up" in the functor lexicon or go through the more complex lexical search (this point is discussed in Section 5).[8] Second, it is hypothesized that the processing that is done with the stressed syllable is, in a sense, "more demanding" than the analysis that is taking place concurrently on the weak syllables. This is because stressed syllable processing involves lexical search, with all the complexities that are linked to it (such as narrowing-in on the appropriate candidate), whereas weak syllable analysis does not.

To illustrate our proposal, we will consider the hypothetical processing of the sentence:

(4) I put the bun in a bag

w s w s w w s
/aɪ pʊt ðə bʌn ɪn ə bæg/

As soon as the listener perceives the first syllable /aɪ/, some analysis is done on it immediately. It may not be identified at this point as the various sources of information do not yet have much to offer. As this analysis is taking place, the system finds the first stressed syllable /pʊt/ and uses it to initiate a lexical search which activates all the words that have /pʊt/ as a stressed syllable. The search comes up with a number of candidates but none that begin with /aɪpʊt/

[8]If function words are listed separately in the lexicon, or in a functor lexicon, in addition to also being listed in the general lexicon, then they do not have to run against non-function word candidates in any search process. This is one of several ways to operationalize the notion of "directly recognizing" function words. It is interesting to add that researchers on discourse (Chafe, 1980, for example) have argued for some form of buffer that contains words (or concepts) that are in focus, that have recently been used, or that constitute common knowledge that is currently in the foreground of attention. When such items are mentioned, they usually pronominalize or delete altogether, but when they do not, they carry little or no stress and low pitch (as in Ladd's, 1980, examples of "destressing"); see for example, Clark and Haviland (1977), Chafe (1980, 1984), Heim (1982), and Kamp (1982). Thus, in addition or along with a function word lexicon, there may be a discourse buffer containing currently focused words. When such words are used, they do not have to compete with the whole lexicon for access, but only with the other words on the list.

and this information is used by the system along with its linguistic knowledge and pattern recognition information to segment /aɪ/ from /pʊt/ and to identify /aɪ/ as the pronoun "I". Information from the lexicon on words that begin with /pʊt/ interacts with the bottom-up, acoustic-phonetic, information as well as phonological, syntactic and semantic knowledge to segment /pʊt/ from /ðə/: no content word in English corresponds to /pʊtðə/, /pʊt/ is a perfectly good word in the growing context, /ðə/ is a very good continuation, etc. The system also knows that no polysyllabic word starts with /ðə/ and that it is therefore most probably the definite article "the". The system finds the next stressed syllable, /bʌn/, and initiates a second lexical search which will use all sources of information to isolate a word that corresponds to the appropriate form class and context. Because the syllable after /bʌn/ is /ɪn/, the system considers words like "bun", but also "bonnet" and "bunny" before appropriately segmenting "bun" and "in" when the nasal information in "in" reduces the possibilities. All this gives the system more syntactic, semantic, pragmatic and prosodic information to help predict and analyze the upcoming phonetic information. The /ə/ before /bæg/ is probably identified and segmented off before the stressed syllable is reached (based on phonotactic constraints, syntactic rules, etc.) but its status as an indefinite article is only confirmed when information comes back from the lexicon: /əbæg/ is not a content word and therefore the listener must be dealing with "a" and "bag".

As noted above, both Cutler (1976) and Bradley (1980) have made very similar proposals to ours. Cutler (1976) suggests that when a stressed syllable is identified, the sentence processor will locate in the lexicon the word of which it is a part, using information about the syllable itself but also about the unstressed syllables immediately before and after it. A number of possible matches may then be found and compared with the (incomplete) information about the unstressed syllables, and this before a choice is made. Once the match is decided upon, the word boundaries can be drawn and the preceding and succeeding words looked up. Cutler goes on to say that since some of the unstressed portions may themselves be words, it follows that words may not necessarily be processed in a left-to-right order. Bradley (1980) has made a very similar proposal. In it she claims that the stressed syllable is picked out of the speech stream and is used to search the lexicon. The candidates are selected on the basis of this syllable and are evaluated by how well they fit with the elements that appear to their left and right.

It is interesting to note that this view of word recognition makes certain predictions about the time it will take to monitor for words in various environments. For example, subjects should be very rapid at monitoring for functions words that do not make up a possible word with the preceding content word (e.g., "him" in "asked him"); they should be much slower when the two make up a possible word ("her" in "got her", in certain dialects, at least), and they should be slower still when the two words make up an actual word, but not the one being monitored for (e.g., "her" in "send her", which may be heard as "center" or "sender"). Some very preliminary data we have obtained would seem to show that such predictions may well be correct.

These first thoughts on a different view of lexical access are of course quite sketchy and general. We do not, for instance, detail how actual content word recognition takes place at the time the lexical search is initiated with the stressed syllable. This is because current models have emphasized this very point and much is known about content word narrowing-in (see Grosjean, 1980; Marslen-Wilson & Welsh, 1978; Nooteboom, 1981; Salasoo & Pisoni, 1985; Tyler & Wessels, 1983 for instance). Rather, we wish to emphasize those points that appear newer to us. First, lexical access is not a WD word by WD word, strictly left to right process (although it might appear to be under certain conditions). Rather, it is a process that uses the prosodic structure of the speech stream and puts particular emphasis on stressed syllables. (Note that this may be a useful approach both in analytic and polysynthetic languages.) Second, two types of analysis appear to take place—a lexical search using the stressed syllable, and a weak syllable analysis. These analyses interact constantly with one another and with the other sources of information (grammatical, situational, etc.) to help segment the speech stream. Third, the system is very much a feed-forward, feed-back system, where there are constant adjustments being made to early and/or partial analyses and constant prediction being made on what is to come.

Although we have little direct evidence for the view we propose, it is interesting to note that research both on speech production and speech perception has constantly emphasized the importance of the stressed syllable. This is precisely what we would expect if our view has any reality whatsoever. Phonetic analyses have shown the stressed syllable to be longer than unstressed syllables, and to have higher pitch and greater amplitude (Lehiste, 1970; Umeda, 1977). The perceptual consequence of this is that stressed syllables are more easily perceived in noise (Kozhevnikov & Chistovitch, 1965, for instance) and may override segmental cues in word identification. It has also been found that the reaction times to the first phoneme of stressed syllables are significantly faster than to phonemes similarly located in unstressed syllables (Cutler & Foss, 1977). Also, mispronunciations are detected more often when the altered segment occurs in stressed syllables than in unstressed syllables (Cole & Jakimik, 1978). And because the stressed syllables of a sentence are component parts of a hierarchical prosodic structure (Liberman & Prince, 1977), their temporal location can be predicted in advance, and this in turn is reflected in various monitoring tasks (Cutler, 1976; Shields, McHugh, &

Martin, 1974). We should note also that slips of the ear very rarely contain mistakes involving stressed syllables (in particular, the vowel of the syllable is seldom wrong) and this has led researchers (Bond & Garnes, 1980, for instance) to propose speech perception heuristics of the type: "Pay attention to stress"; "Pay attention to stressed vowels", etc.). Finally, Huttenlocher and Zue (1983) have found that when words are classified according to their stress pattern and to a broad phonetic transcription of their stressed syllable only, then 17% of these words are uniquely specified in the lexicon, 38% belong to a class of 5 or less and the average class size is 3.8. We can conclude from this that the stress pattern of a word, and the phonetics of its stressed syllable, are valuable information for the lexical access system; it would be surprising if it did not make use of it!

Thus, past and current research in the production and perception of language points to the stressed syllable as an important unit in processing. Given the evidence that stress is crucial to processing, it is interesting to conjecture why it has been ignored in almost all theories of lexical access. The reason, as we suggested above, is that psycholinguistic theories have almost always been constructed in terms of syntactic units—words and phrases. Once we think in terms of prosodic units, then our notion of the word and the phrase change, as does our view of processing.

5. The function word vs. content word issue

The distinction between function words and content words has played an important role in recent psycholinguistics (see, e.g., Bradley, 1978; Bradley, Garrett, & Zurif, 1979; Friederici & Schoenle, 1980; Garrett, 1975, 1976, 1978, 1980; Segui, Mehler, Frauenfelder, & Morton, 1982; Swinney, Zurif, & Cutler, 1980), and it is therefore important to examine it in light of our view of lexical access. We wish to argue that the distinction between the two types of words is not as relevant during lexical access as was once thought; rather a function word will be processed differently depending on its information load, its degree of stress, its length, and so on. The more salient a function word becomes (through slower and more careful pronunciation or through increased stress), the more it will be accessed like a content word. It is not that function words are processed one way and content words another, but rather that stressed and unstressed syllables are processed differently. In fact, Cutler and Foss (1977) have shown evidence for this in a phoneme monitoring task: they found no reaction time differences between function and content words when stress was held constant, showing thereby that form class itself is not the critical factor; it is the stress on the item that is. Note also that Kean (1979) makes precisely the same arguments in her work on aphasia.[9]

Some pilot data we have obtained appear to show that there is a strong negative correlation between the duration of a monosyllabic function word and its monitoring time: long functions words (such as "some") take less time to monitor than short function words (such as "a" and "of"). But duration is in part dependent on the information load carried by a word; the more important the content of the word and the less redundant its information in context, the more salient it will be at the prosodic level (in terms of pitch, amplitude and duration), and hence the more quickly it will be monitored. Thus, the information load of a word in various contexts will determine in large part its saliency. In turn this saliency will determine which of two sorts of analyses it falls under: whether it will be accessed in depth through the mental lexicon, or whether it will be subject to a weak syllable analysis. This implies that in a rich enough context, even content words might be redundant enough to lose saliency, and in other sorts of context, function words might become informationally important, bear stress, and be accessed through the main lexicon.

It is important to note, therefore, that there is no clear two-way distinction between function words and content words, but rather a continuum of cases ranging from forms with full contentful meaning that bear a lot of stress, to forms that carry less content but more grammatical meaning, and thus bear less stress. Forms towards the more contentful end of the continuum tend to retain more phonological identity in the speech stream, while forms towards the grammatical end tend to retain less phonological identity and to contract tighter bonds with the stressed forms they are phonologically linked to. Table 1 presents a continuum of forms that range from those that carry more information, more stress, and retain more phonological identity, to forms that carry less information, less stress and retain less phonological identity. The lower the form on the list, the more likely it is to "merge" with a form that carries stress. Although the table represents the unmarked cases, it is clear that particular contexts may render a more contentful form redundant, and a more grammatical form informationally important, thus changing their prosodic characteristics.

[9]It should also be noted that function words sometimes have different functions and meanings when they are stressed. For example, the differences between stressed and unstressed "some" and "many" (often represented as "sm" versus "some" by linguists) is well known. "Sm" is an indefinite plural article whereas "some" is usually a restricted existential quantifier (for differences in their behavior, and other such pairs, see Milsark, 1977).

There appears to be no principled way, therefore, to break the continuum into categories so as to obtain two classes of "words": function and content. It is our contention that a view of the lexical access process based on the prosodic structure of the speech stream can best account for the full range of cases in the table. They will fall under one or the other of our two types of analysis (and even under both at times) depending on the context they are in. The important distinction is not one between content and function words per se, but one between salient (strong) versus non-salient (weak) syllables and the two types of analysis they involve.

Table 1. *A continuum of cases going from forms that typically carry more information, more stress, and retain more phonological identity to forms that carry less specific information, less stress, and retain less phonological identity. The lower a form on the list, the more likely it is to merge, in a more or less tight phonological bond, with a form that carries stress. However, in context, any form can be destressed or stressed*

1.	Content words carrying new information in context
2.	Content words carrying old information in context (destressed)
3.	Numerals, quantifiers, adverbs
4.	Stressed "some" and "many"
5.	Unstressed "some" and "many"
6.	Unreduced auxiliary verbs
7.	Full pronouns
8.	Reduced auxiliaries
9.	Reduced pronouns
10.	Bisyllabic prepositions
11.	Content prepositions
12.	Grammatical prepositions (case markers)
13.	Demonstrative articles
14.	The definite article "the"
15.	The indefinite article "a(n)"
16.	The infinitive marker "to"
17.	The syntactically determined preposition "of"
18.	Contracted forms, like "hadda" or "fronta"
19.	Prefixes/suffixes
20.	Inflectional morphemes
21.	Secondary stressed syllables of polysyllabic words
22.	Syllables with unreduced vowels in polysyllabic words
23.	Unstressed syllables in polysyllabic words

6. Concluding remarks

The basic aim of our paper was to call attention to the important role that the prosodic structure of spoken language may play during the recognition of words in continuous speech. The strong influence of written language on current views of spoken language processing has led to the development of models of lexical access that are giving too much importance to the written notion of the word. Recent developments in prosodic theory encourage us to pay attention to prosodic units and to the alternation of weak and strong syllables in the speech stream. Our view of continuous spoken word recognition is an attempt to integrate the prosodic structure of language into a processing model. In addition, it helps explain why words are not all recognized sequentially, one after the other, and it has interesting implications for the controversy that surrounds the function versus content word issue. Our view may well be wrong, but it will have served its purpose if it forces us to take into account prosodic structure in our models of spoken language processing.

References

Bickerton, D. (1981). *Roots of language*. Ann Arbor: Karoma Publishers.

Bolinger, D. (1958a). A theory of pitch accent in English. *Word, 14*, 109–149.

Bolinger, D. (1958b). Stress and information. *American Speech, 33*, 5–20.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language, 37*, 83–69.

Bolinger, D. (1965). *Forms of English: Accent, morpheme, order*. Cambridge, MA: Harvard University Press.

Bolinger, D. (1972). Accent is predictable (if you're a mind reader). *Language, 48*, 633–644.

Bolinger, D. (1981). Two kinds of vowels, two kinds of rhythm. Bloomington, Indiana: Indiana University Linguistics Club.

Bond, Z.S., & Garnes, S. (1980). Misperceptions of fluent speech. In R. Cole (Ed.) *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

Bradley, D. (1978). *Computational distinctions of vocabulary type*. Unpublished doctoral thesis, MIT.

Bradley, D. (1980). Lexical representation of derivational relation. In M. Aronoff & M.-L. Kean (Eds.) *Juncture*. Saratoga, CA: Anma Libri.

Bradley, D., Garrett, M.F., & Zurif, E.B. (1979). Syntactic deficits in Broca's aphasics. In D. Caplan (Ed.) *Biological studies of mental processes*. Cambridge, Mass.: MIT Press.

Chafe, W.L. (1980). The deployment of consciousness in the production of a narrative. In W.L. Chafe (Ed.) *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

Chafe, W. (1984). Cognitive constraints on information flow. *Berkeley Cognitive Science Report No. 26*. Berkeley Cognitive Science Program, Institute of Cognitive Studies, University of California at Berkeley.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Clark, H.H., & Haviland, S.E. (1977). Comprehension and the given-new contract. In R.O. Freedle (Ed.) *Discourse production and comprehension*. Norwoord, NJ: Ablex.

Cole, R.A., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.) *Strategies of information processing*. New York: Academic Press.

Cole, R.A., & Jakimik, J. (1979). A model of speech perception. In R. Cole (Ed.) *Perception and production of fluent speech.* Hillsdale, NJ: Erlbaum.

Comrie, B. (1981). *Language universals and linguistic typology.* Chicago: University of Chicago Press.

Cotton, S., & Grosjean, F. (1984). The gating paradigm: A comparison of successive and individual presentation formats. *Perception and Psychophysics, 35*,41–48.

Culicover, P., & Rochemont, M. (1983). Stress and focus in English. *Language, 59*, 123–165.

Cutler, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics, 20*, 55–60.

Cutler, A., & Foss, D.J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20*, 1–10.

Forster, K. (1976). Accessing the mental lexicon. In R. Wales, & E. Walker (Eds.) *New approaches to language mechanisms.* Amsterdam: North-Holland.

Frazier, Lyn (1987). Structure in auditory word recognition. *Cognition, 25*, 157–187.

Friederici, A.D., & Schoenle, P.W. (1980). Computational dissociation of two vocabulary types: evidence from aphasia. *Neuropsychologia, 18*, 11–20.

Fries, C.C. (1952). *The structure of English: An introduction to the construction of English sentences.* New York: Harcourt, Brace.

Garrett, M.F. (1975). The analysis of sentence production. In G. Bower (Ed.) *Psychology of learning and motivation.* New York: Academic Press.

Garrett, M.F. (1976). Syntactic processes in sentence production. In R.J. Wales, & E. Walker (Eds.) *New approaches to language mechanisms.* Amsterdam: North-Holland.

Garrett, M.F. (1978). Word and sentence perception. In R. Held, H.W. Leibowitz, & H.L. Teuber (Eds.) *Handbook of sensory physiology, Vol. VIII: Perception.* Berlin: Springer-Verlag.

Garrett, M.F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.) *Sentence production.* London: Academic Press.

Garrett, M.F., & Kean, M.-L. (1980). Levels of representation and the analysis of speech errors. In M. Aronoff, & M.-L. Kean (Eds.) *Juncture.* Saratoga, Ca.: Anma Libri.

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

Givon, T. (1975). Serial verbs and syntactic change: Niger-Congo. In C.N. Li (Ed.) *Word order and word order change.* Austin/London: University of Texas Press.

Givon, T. (1979). *On understanding grammar.* New York: Academic Press.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28*(4), 267–283.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception and Psychophysics, 38*, 299–310.

Halle, M., & Keyser, S.J. (1971). *English stress: Its form, its growth, and its role in verse.* New York: Harper and Row.

Hayes, B. (1980). *A metrical theory of stress rules.* Unpublished Ph.D. Dissertation, MIT.

Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry, 13*, 227–276.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases.* Doctoral dissertation, University of Massachusetts, Amherst.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America, 29*, 296–305.

Huttenlocher, D., & Zue, V. (1983). Phonotactic and lexical constraints in speech recognition. *Working Papers of the Speech Communication Group at MIT, 3*, 157–167.

Jakimik, J. (1979). Word recognition and the lexicon. In J. Wolf, & D. Klatt (Eds.) *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America.* New York: Acoustical Society of America.

Kamp, H. (1982). A theory of truth and semantic representation. In J. Groenendijk et al. (Eds.) *Formal methods in the study of language*. Amsterdam: North-Holland.
Kean, M.-L. (1979). Agrammatism: a phonological deficit? *Cognition*, *7*, 69–83.
Kean, M.-L. (1980). Grammatical representations and the description of processing. In D. Caplan (Ed.) *Biological studies of mental processes*. Cambridge, Mass.: MIT Press.
Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry*, *10*, 421–442.
Klatt, D. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*, 279–312.
Kozhevnikov, V., & Chistovich, L. (1965). *Speech: Articulation and perception*. US Department of Commerce Translation, IPRS 30, 543, Washington, DC.
Ladd, D.R. (1980). *The structure of intonational meaning*. Bloomington: Indiana University Press.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass.: MIT Press.
Liberman, M. (1975). *The intonational system of English*. Unpublished Ph.D. Dissertation, MIT.
Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336.
Marslen-Wilson, W. (1984). Function and process in spoken word recognition. In H. Bouma, & D. Bouwhuis (Eds.) *Attention and performance: Control of language processes*. Hillsdale, NJ: Erlbaum.
Marslen-Wilson, W., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.
Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.
McCarthy, J. (1979). On stress and syllabification. *Linguistic Inquiry*, *10*, 443–466.
McClelland, J., & Elman, J. (1986). The Trace Model of Speech Perception. *Cognitive Psychology*, *18*, 1–86.
Mehler, J., Segui, J., & Carey, P. (1968). Tails of words: Monitoring ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *17*, 29–35.
Milsark, G.L. (1977). Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis*, *3*, 1–29.
Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165–178.
Morton, J., & Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, *15*, 43–51.
Nespor, M., & Vogel, I. (1979). Clash avoidance in Italian. *Linguistic Inquiry*, *10*, 467–482.
Nespor, M., & Vogel, I. (1982). Prosodic domains of external sandhi rules. In H. van der Hulst, & N. Smith (Eds.) *The structure of phonological representations (Part I)*. Dordrecht: Foris.
Nespor, M., & Vogel, I. (1983). Prosodic levels above the word and ambiguity. In A. Cutler, & D.R. Ladd (Eds.) *Prosody: Models and measurements*. Berlin/Heidelberg/New York: Springer-Verlag.
Nooteboom, S. (1981). Lexical retrieval from fragments of spoken words: Beginnings and endings. *Journal of Phonetics*, *9*, 407–424.
Pollack, I., & Pickett, J.M. (1963). The intelligibility of excerpts from conversation. *Language and Speech*, *6*, 165–171.
Pollack, I., & Pickett, J.M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, *3*, 79–84.
Prince, A. (1983). Relating to the grid. *Linguistic Inquiry*, *14*, 19–100.
Rubenstein, H., & Pollack, I. (1963). Word predictability and intelligibility. *Journal of Verbal Learning and Verbal Behavior*, *2*, 147–158.
Safir, K. (Ed.) (1979). Papers on syllable structure, metrical structure, and harmony processes. MIT Working Papers in Linguistics 1.
Salasoo, A., & Pisoni, D. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Cognition*, *2*, 210–231.

Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia, 20*, 615–627.
Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. Paper presented at the Conference on the Mental Representation of Phonology, University of Massachusetts at Amherst, 1978. In T. Fretheim (Ed.) *Nordic Prosody II*. Trondheim: TAPIR.
Selkirk, E. (1980a). The role of prosodic categories in English word stress. *Linguistic Inquiry, 11*, 563–605.
Selkirk, E. (1980b). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff, & M.-L. Kean (Eds.) *Juncture*. Saratoga, Ca.: Anma Libri.
Selkirk, E.O. (1982). *The syntax of words. Linguistic Inquiry Monograph* 7. Cambridge, MA.: MIT Press.
Selkirk, E.O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA.: MIT Press.
Shields, J.L., McHugh, A., & Martin, J.G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology, 102*, 250–255.
Slobin, D.I. (1977). Language change in childhood and history. In J. Macnamara (Ed.) *Language learning and thought.* New York: Academic Press.
Slobin, D.I. (1982). Universal and particular in the acquisition of language. In E. Wanner, & L.R. Gleitman (Eds.) *Language acquisition: The state of the art.* Cambridge: Cambridge University Press.
Swinney, D. (1982). The structure and time-course of information interaction during speech comprehension: Lexical segmentation, access and interpretation. In J. Mehler, E. Walker, & M. Garrett (Eds.) *Perspectives on mental representation*. Hillsdale, NJ: Erlbaum.
Swinney, D., Zurif, E.B., & Cutler, A. (1980). Effects of sentential stress and word class upon comprehension in Broca's aphasics. *Brain and Language, 10*, 132–144.
Taft, L. (forthcoming). *Prosodic constraints and lexical parsing strategies.* Doctoral dissertation, University of Massachusetts, Amherst.
Tyler, L.K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics, 34*(5), 409–420.
Tyler, L.K., & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception and Psychophysics, 38*, 217–222.
Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America, 61*, 846–858.
Williams, E. (1981). On the notions "lexically related" and "head of a word." *Linguistic Inquiry, 12*, 245–274.
Zwicky, A.M. (1977). On clitics. Paper presented at the 3rd International Phonologie-Tagung at the University of Vienna, 1976. Bloomington, Indiana: Indiana University Linguistics Club.
Zwicky, A.M. (1982). Stranded *to* and phonological phrasing in English. *Linguistics, 20*, 3–57.

Résumé

Le but de cet article est d'attirer l'attention sur le rôle que joue la structure prosodique dans la reconnaissance des mots. Nous commençons par soutenir que la notion de mot écrit a eu une influence bien trop grande sur les modèles de la reconnaissance des mots parlés. Nous discutons ensuite plusieurs propriétés de la structure prosodique qui sont importantes pour les questions de reconnaissance. Nous présentons ensuite une conception de la reconnaissance des mots "en continu" qui tient compte de l'alternance entre syllabes fortes et faibles dans le flux de la parole. L'Accès lexical est guidé par les syllabes fortes, tandis que les syllabes faibles sont identifiées par une analyse globale de leur distribution et par l'utilisation de règles phonotactiques et morphonémiques. Nous concluons par une discussion de la controverse sur les différences d'accès aux mots à contenu et aux mots à fonction à la lumière de notre conception.