*Prosody in the Comprehension of Spoken Language: A Literature Review**

ANNE CUTLER
DELPHINE OAHAN
WILMA van DONSELAAR

Max-Planck-Institute for Psycholinguistics

KEY WORDS

accent

comprehension of spoken language

prosody

stress

suprasegmental structure

ABSTRACT

Research on the exploitation of prosodic information in the comprehension of spoken language is reviewed. The research falls into three main areas: the use of prosody in the recognition of spoken words, in which most attention has been paid to the question of whether the prosodic structure of a word plays a role in initial activation of stored lexical representations; the use of prosody in the computation of syntactic structure, in which the resolution of global and local ambiguities has formed the central focus; and the role of prosody in the processing of discourse structure, in which there has been a preponderance of work on the contribution of accentuation and deaccentuation to integration of concepts with an existing discourse model. The review reveals that in each area progress has been made towards new conceptions of prosody's role in processing, and in particular this has involved abandonment of previously held deterministic views of the relationship between prosodic structure and other aspects of linguistic structure.

PROSODY AND PROCESSING

Prosody is an intrinsic determinant of the form of spoken language. The prosodic structure of an utterance exercises effects on the timing, amplitude, and frequency spectrum of the utterance, and these are the dimensions of sound itself; any utterance, indeed any part of an utterance corresponding to any linguistic component, to a phonetic segment even, must have a certain duration, a certain amplitude, a certain fundamental frequency (see Lehiste, 1970, for elaboration of this point). Whenever listeners recognize normal speech, they are processing prosodically determined variation.

* Acknowledgements: We thank the editors *of Language and Speech* for the invitation to pool our expertise in the form of this review, and Mary Beckman, Julia Hirschberg, James McQueen, Peter Roach, Mark Steedman, and Jacques Terken for their helpful advice on an earlier version of the text. The second author acknowledges support from the Fyssen Foundation, Paris. The current address of the second author is: Dept of Cognitive Science, Johns Hopkins University, Baltimore MD 21218, U.S.A.

Address for correspondence: Anne Cutler, Max-Planck-Institute for Psycholinguistics, P O. Box 310, 6500 AH Nijmegen, The Netherlands. Tel. +31 24 352 1911. E-mail: anne.cutler@mpi.nl

This might seem to suggest that all research on the recognition of speech (possibly with the exception of certain machine-generated or disturbed speech) could somehow be considered relevant to the topic of the present review. But readers will not, we assume, expect to find under this title a complete overview of all psycholinguistic studies of the comprehension of speech. We attempt to summarize here the research that readers will expect us to cover; but because this is an interdisciplinary journal, expectations may vary. The term *prosody* is used in different ways by different researchers within the *Language and Speech* fold: from at one extreme those who maintain an abstract definition not necessarily coupled to any statement about realization ("the structure that organizes sound"), to those who use the term to refer to the realization itself, that is, effectively use it as a synonym for *suprasegmentalfeatures* ("pitch, tempo, loudness, pause") at the other extreme. It would surprise the latter group that the former would consider the structure of syllables to fall within the study of prosody, and it would surprise the former group that the latter group would similarly include questions of speaker identification. Perhaps the majority of readers would fall between these extremes, in using the term to refer to abstract structure coupled to a particular type of realization ("the linguistic structure which determines the suprasegmental properties of utterances"); but there is a continuum of approaches, and no one definition is valid for all the research that we review. We have chosen to be comprehensive rather than restrictive in our selection (if only because a comprehensive review can best help to prevent wasted repetition of effort).

Nevertheless, the research we review is unified by its aim: to understand the process of human recognition of spoken language. The task of the listener is to reconstruct the speaker's message, and there are various different aspects to this task: recognizing the individual words, extracting their syntactic relationships, determining the semantic structure of the utterance and its relation to the discourse context. There is a body of research—much of it quite recent, as steadily more psycholinguistic work has concerned spoken language—on each of these topics: the role of prosodic structure in lexical recognition, in syntactic processing, and in the comprehension of discourse structure. The three central sections of our review, below, deal with these issues in turn. Aspects of the listener's recognition performance which we have excluded from this review, but which may be considered related in so far as they involve the processing of suprasegmental information, include phonetic segment identification (research which falls traditionally under the label of phonetics rather than Psycholinguistics); speaker identification and recognition of dialect; and recognition of emotion. And because we are concerned with research on human recognition only, we do not cover any of the research involving the use of prosodic information in speech recognition by machine (see Waibel, 1988; Price & Ostendorf, 1996; Sagisaka, Campbell, & Higuchi, 1997, for overviews).

The methods used in the greater part of the research we review are those of current Psycholinguistics: measurements of response time to perform some task such as detection of a target or decision about the lexical status of a word; question-answering tasks, ratings of appropriateness, judgments of familiarity, and other methods for assessing ease and outcome of processing. Probably the majority of studies attempt to come to grips with the processing (albeit under controlled conditions) of intact prosody; but there is also a line of research which investigates the role of prosody in processing via observation of the effects of ill-formed prosody (e.g. mis-stressed words, mismatch between sentence accent and

focus structure). The aim in all cases is to understand (and model) the recognition process. We do not include research which is primarily aimed at an understanding of prosodic structure itself.

This unity of research aim however does not guarantee unity in underlying assumptions. Recent developments in research on prosody have brought closer together research traditions which had initially developed separately. A decade and a half ago, Ladd and Cutler (1983) drew attention to two traditions in the study of prosody, which differed both in their methodology (instrumental/experimental vs. theoretical/descriptive) and in their assumptions about the structure of prosody (concrete vs. abstract). Such a division no longer accurately characterizes the field. Two main developments seem to be responsible for this change. First, many findings showed that prosody did not directly reflect the grammar—that is, syntactic structure could not be directly mapped onto acoustic features. As a result, researchers became willing to entertain more complex accounts of prosodic structure. Second, researchers within the theoretical tradition felt a need to bolster their proposals with phonetic evidence—that is, they altered their traditional methodology. As will be described below when we discuss the computation of syntax, many recent studies of prosodic processing — in speech perception, speech production, and language acquisition—have used a theoretical model as a framework, and have attempted to describe the links between this abstract representation and the linguistic function or representation under investigation (e.g., Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Wightman, Shattuck-Hufhagel, Ostendorf, & Price, 1992; Warren, Grabe, & Nolan, 1995; Ferreira, 1993; Gerken, 1994; 1996). At the same time, however, other researchers have addressed similar processing issues from the perspective of acoustic form (Beach, 1991; Pitt & Samuel, 1990; Sanderman, 1996; Stirling & Wales, 1996). We therefore review a body of work whose methodology is relatively unified but whose theoretical assumptions, as we will attempt to make clear, can vary.

Quite early in the history of Psycholinguistics, research findings suggested that the prosodic structure of an utterance plays an organizing role in speech recognition. A classic finding by Epstein (1961) that a string of nonsense syllables is recalled better if presented with sentence morphology *(meeving gups keebed gompily)* than without *{meev gup keeb gomp)* was demonstrated to hold under auditory presentation only if the strings were also spoken with sentence prosody; there was no facilitation if the inflected words were read as a list (Leonard, 1974; O'Connell, Turner, & Onuska, 1968). Similarly, although grammatical strings are more easily shadowed than ungrammatical strings (Miller & Isard, 1963), the superiority of grammatical strings disappears if they do not also have sentence prosody (Martin, 1968). Zurif and Mendelsohn (1972) even found that the right-ear advantage, which is found for the auditory perception of speech, was greater for nonsense strings read as a sentence than as a list. Further, the prosodic structure of a heard utterance forms part of the memory representation which listeners form of the input. Robinson (1977) presented listeners with nonsense bisyllables varying in stress pattern (e.g., *BIsev, juBIM, JUlem*—note that throughout this review we will use upper case to signify accent) and found that false alarms in a recognition test were more likely to be made to foil-items which preserved the stress of individual syllables from the original presentation (e.g., *BIlem,* when *BIsev* and *JUlem* had been presented) than to foil-items which changed a syllable's stress (e.g., *biLEM).* Speer, Crowder, and Thomas (1993) found that previously heard

sentences, and even nonsense utterances, could be recognized more accurately on a second presentation if they were spoken with the same prosody as on their first presentation. Listeners' attention to the organizing function of speech prosody was highlighted by the finding of Darwin (1975) that when prosodic continuity and semantic continuity conflict, listeners attend to the former; listeners whose task is to repeat one of two simultaneous messages presented one to each ear will switch their shadowing to the wrong channel to follow prosodic continuity, even if the effect of the switch is that the message they are repeating becomes nonsensical. Similarly, synthesized speech, often judged to be deficient in prosodic naturalness, is more likely to receive listener approval when the prosody is improved (Silverman, 1987; Terken & Lemeer, 1988), and this may also result in more efficient processing (Larkey & Danly, 1983; Sanderman, 1996; Silverman, Kalyanswamy, Silverman, Basson, & Yashchin, 1993; Sorin & Le Bras, 1983). The converse of the facilitation effect of consistent prosody is deterioration of processing efficiency when prosody is disrupted. Thus disruption of prosodic structure via cross-splicing of utterances or insertion of silent intervals leads to longer latencies in word or phoneme target detection tasks (Buxton, 1983; Martin, 1979; Meltzer, Martin, Mills, Imhoff, & Zohar, 1976; Tyler & Warren, 1987); disruption via fragmented presentation of an utterance impairs listeners' ability to distinguish the intended interpretation of an ambiguous sentence (Ferreira, Anes, & Horine, 1996).

All these lines of research combine to show that processing of speech input is facilitated in several ways by coherent prosodic structure appropriate for sentences. More recent investigations of such facilitatory effects have established a crucial role for temporal patterning. Thus temporal envelopes of spoken utterances, preserving amplitude information but virtually without spectral variation, allow listeners to recognize short utterances and even nonsense syllables almost perfectly (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Likewise, sinusoidal replicas of natural utterances are readily intelligible (Remez, Rubin, Pisoni, & Carrell, 1981) and even convey a good deal of information about phonetic structure and speaker identity (Remez, Fellowes, & Rubin, 1997; Fellowes, Remez, & Rubin, in press). Importantly, this "sinewave speech" is not perceived as speech if the signal it replicates consists of a sequence of vowels (Remez et al., 1981), indicating that the temporal patterning is critical. So far, there has been no attempt to fill in the gap between these findings and the earlier observations of facilitation from sentence-level prosodic structure, but there are clearly research questions which arise: is such a transformation of an utterance more likely to be, or more easily, processed as speech if the utterance has intact as opposed to disrupted prosodic structure?

It is clear from these general findings that listeners attend to the prosodic structure of heard utterances. In the following sections we consider the question of how prosodic structure is specifically exploited in spoken-language understanding, respectively in the recognition of lexical items, of syntactic structure, and of discourse relationships.

PROSODY IN THE RECOGNITION OF SPOKEN WORDS

Introduction

Current psycholinguistic models of spoken-word recognition conceive of this process as one of activation of, and competition between, potential candidate words (see Cutler, 1995, for

a review). Where might prosodic information play a role in word recognition? Firstly, we can ask whether listeners use prosodic information to activate the stored lexical representations of the individual words in the utterance. In many languages, words may be distinguished solely by differences in internal prosodic structure, and such differences may constrain even the earliest stages of lexical activation. But it is also possible that prosody is irrelevant to activation but may still serve to select among alternative candidate words — for example, if the competition process fails to produce a clear winner. Further complicating the issue is the fact that words are rarely spoken in isolation; they occur within a surrounding context of (more or less) fluent speech. Thus we can also ask whether prosodic information in an utterance serves, even in the absence of lexically distinctive prosody, to provide the listener with information as to the number and location of individual words—that is, lexical units which have to be separately recognized in order for the utterance as a whole to be understood. These two topics are not independent, although they may be said to have an in-built logical order in that any contribution of prosodic processing to word-boundary location, and hence to initiation of lexical access, should operate prior to any role played by prosody in the lexical access process itself. Indeed, there is evidence that prosodic structure in English does exercise prior constraints on the activation and competition process (McQueen, Norris, & Cutler, 1994). The question of word-boundary location in continuous speech has therefore been seen by psycholinguists as part of the prelexical processing of speech input.

We therefore begin this part of our review by considering prelexical processing. The following subsection summarizes empirical evidence on the role of various forms of lexical prosody in the process of spoken-word recognition.

Prosody and prelexical processing

In most spoken language, few cues are available to signal exactly where one word ends and the next begins. However, the comprehension of spoken language must involve processing discrete words, rather than utterances as indivisible wholes, because most complete utterances have never previously been experienced by the listeners to whom they are directed. To understand an utterance, therefore, listeners must somehow, very often in the absence of explicit signals, locate the boundaries between the individual words (or more precisely, the lexically represented units) of which the utterance is composed.

Studies of word boundary perception have in general given more attention to allophonic cues than to prosodic variation, but at least some have considered relative timing as a segmentation cue. Nakatani and Schaffer (1978) found that listeners performed significantly better than chance at distinguishing English adjective-noun sequences such as *noisy dog* and *bold design* when these were presented as reiterant speech (in which a natural utterance is mimicked in a series of repetitions of a single syllable such as "ma"), and that the most effective cue was the relative syllable duration of the phrases. Similar results for French ambiguous strings *(le couplet complet - le couple est complet)* were reported by Rietveld (1980). Segmental-level timing was also shown to be important in a series of studies by Quene (1987, 1989, 1992, 1993) investigating minimal junctural pairs in Dutch; he found that speakers frequently differed considerably in how they signaled word boundaries, but that listeners could make use of whatever cues speakers provided. The most accurately located word boundaries were those falling after a consonant and before a vowel; in this context, the overall duration of the consonant was a rather stronger cue than the time taken by the

vowel to reach its maximum amplitude, and sonorant consonants provided better prejunctural cues than fricatives or plosives.

Many studies of English have shown how segmental duration varies with position in a word—consonants, for example, tend to be longest in word-initial position, somewhat shorter in word-final position, and shortest in word-medial position (Oiler, 1973; Klatt, 1974, 1976; Umeda, 1977; but see van Son & van Santen, 1997). Vowel duration on the other hand is primarily determined by stress rather than position in the word (Umeda, 1975; Crystal & House, 1988, 1990). Of course, syllable duration also varies with position in the word; if stress and syllable weight are controlled, word-final syllables are somewhat longer than nonfinal syllables (Oiler, 1973; Klatt, 1975), though some investigators have failed to find such effects (Harris & Umeda, 1974). Klatt (1976) speculated that durational cues to word-boundary location in English are too small and too variable to be of much perceptual use.

A series of studies in recent years has motivated an explanation of the lexical segmentation component of human sentence processing in what may be argued to be prosodically based terms. First, experiments carried out in English suggested that listeners segment speech at the onset of strong syllables (syllables with full vowels that can potentially be stressed). For example, finding a real word in a spoken nonsense sequence is hard if the word is spread over two strong syllables (e.g., *mint* in [mintef]) but easier if the word is spread over a strong and a following weak syllable (e.g. *mint* in [mintef]; Cutler & Norris, 1988). The proposed explanation for this is that listeners divide the former sequence at the onset of the second strong syllable, so that detecting the embedded word requires recombination of speech material across a segmentation point; the latter sequence offers no such obstacles to embedded-word detection, as the noninitial syllable is weak and so the sequence is simply not divided. Similarly, when English-speakers make slips of the ear which involve mistakes in word boundary placement, they tend most often to insert boundaries before strong syllables (e.g., hearing *by loose analogy* as *by Luce and Allergy)* or delete boundaries before weak syllables (e.g., hearing *how big is it?* as *how bigoted?;* Cutler & Butterfield, 1992). The proposal that listeners segment English speech at strong syllable onsets is well justified by distributional patterns in the input (Cutler & Carter, 1987); strong syllables are indeed highly likely to signal the onset of lexical words. Distributional patterns in the Dutch vocabulary are similar to those of English (van Heuven & Hagman, 1988; Schreuder & Baayen, 1994), and results similar to those both of Cutler and Norris (1988) and Cutler and Butterfield (1992) are found with Dutch (Vroomen, van Zon, & de Gelder, 1996).

Neither English nor Dutch has what is known as fixed stress—that is, an obligatory word-internal position for stress (fixed-stress languages include, for example, Finnish, which has stress on the first syllable of every word, or Polish, in which stress always falls on the penultimate syllable). It might be imagined that fixed stress could provide an excellent cue to word-boundary location; to our knowledge, however, no empirical demonstration exists of the exploitation of stress cues in segmentation of such a language. In fact, it is, rather paradoxically, possible that fewer explicit acoustic correlates of stress in fixed-stress languages may be available for listeners' use than in free-stress languages; when stress is fully predictable, its explicit realization may be unnecessary. However, experimental work on this issue would be welcome. French, although not a lexical stress language, has a consistent prosodic pattern which could provide information about some word boundaries:

accent falls on the final syllable of rhythmic groups, and the right boundary of a rhythmic group is always also the right boundary of a word. There is some evidence that French listeners can use this regularity to speed detection of a target syllable located at a rhythmic-group boundary in comparison to the same syllable at another location (Dahan, 1996). Extension of this type of research to fixed-stress languages would provide useful knowledge. Listeners' segmentation of English and Dutch apparently exploits the asymmetries of stress placement in those vocabularies; in the case of English, however, it is best described not as segmentation via the primary stress placement of words (although see Grosjean & Gee, 1987, for a proposal couched in these terms), but rather via the pattern of strong versus weak syllables, or what may be termed the utterance rhythm.

Evidence from Swedish shows that listeners can perceive the pattern of strong and weak syllables without reference to segmental information. Svensson (1971, 1974) found that listeners presented with hummed Swedish speech produced candidate transcriptions which accurately preserved the pattern of strong and weak syllables. Carlson, Granstrom, Lindblom, and Rapp (1972) found that Swedish listeners could accurately reproduce the pattern of strong and weak syllables in reiterant speech; importantly, however, they were not in general accurate in locating the word boundaries.

The opposition between strong and weak syllables is of course an important feature of the phonology of English and prosodically similar languages; but other languages have quite different phonologies. French, for example, does not exhibit such contrasts between strong and weak syllables. Evidence from a wide variety of experimental tasks in French suggests that listeners' processing of spoken language involves a process of segmentation of the input into syllable-sized units. Consider, for instance, the French words *balance* and *balcon;* although they begin with the same three phonemes, the first two phonemes are grouped together by listeners as the initial unit in *ba-lance*, the first three in *hal-con* (Mehler, Dommergues, Frauenfelder, & Segui, 1981; Segui, Frauenfelder, & Mehler, 1981; Cutler, Mehler, Norris, & Segui, 1986; Dupoux & Mehler, 1990; Kolinsky, Morais, & Cluytens, 1995; Pallier, Sebastian-Galles, Felguera, Christophe, & Mehler, 1993; Peretz, Lussier, & Beland, 1996).

The apparent difference between the experimental results for English and French can in fact be viewed as a similarity: both stress in English and the syllable in French are the basis of rhythmic structure in their respective languages. English is the paradigm example of a stress-timed language and French of a syllable-timed one (Pike, 1945; Abercrombie, 1967). This symmetry prompted the hypothesis (see, e.g., Cutler, Mehler, Norris, & Segui, 1992) that listeners might in fact adopt a universally applicable solution to the word-boundary problem in that, to solve it, they exploit whatever rhythmic structure happens to characterize their language; the boundaries of the units of rhythmic structure will in general correspond (although not necessarily in a one-to-one manner) to boundaries between words. This proposal in turn implied that where a language has a rhythmic structure based on some phonological construct other than stress or the syllable, it should be possible to find evidence for exploitation of such rhythmic structure in speech segmentation. Japanese is such a language; its rhythm is described in terms of a subsyllabic unit, the mora (e.g., the word *tanshi* has three morae: *ta-n-shi).* Otake, Hatano, Cutler, and Mehler (1993) and Cutler and Otake (1994) provided evidence favoring mora-based segmentation by Japanese listeners (see also Otake, Yoneyama, Cutler, & van der Lugt, 1996).

Note that the presence of a particular rhythmic structure in the input does not of itself produce segmentation based on that structure. English listeners show no evidence of syllabic segmentation with French input, for example (Cutler et al., 1986), and neither do Japanese listeners (Otake, Hatano, & Yoneyama, 1996); English listeners likewise show no evidence of mora-based segmentation of Japanese input (Otake et al., 1993; Cutler & Otake, 1994), and nor do French listeners (Otake et a1, 1993). The segmentation procedures are therefore not brought into play simply by characteristics of the input; instead, they form part of the processing repertoire of the listener and differ as a function of the listener's language experience. Indeed, given the opportunity, listeners will apply their native language-specific procedures to foreign language input, even in cases where the procedures may not operate efficiently at all. Thus French listeners apply syllabic segmentation to English words such as *balance* and *balcony* where English listeners do not (Cutler et al., 1986); likewise, they apply syllabic segmentation to Japanese input (e.g., preferring to segment *tanshi* as *tan-shi;* Otake et al., 1993); and Japanese listeners apply moraic segmentation where possible to English input (e.g., showing facilitated processing of the syllable-final nasal in words like *fender,* where English listeners do not; Cutler & Otake, 1994).

This body of work thus suggests that prosodic structure, in the form of language rhythm, helps listeners to perform lexical segmentation efficiently. In English, as the previous section showed, the strong/weak syllable distinction is based primarily on a segmental property (vowel quality) rather than on a stress distinction (Fear, Cutler, & Butterfield, 1995). Furthermore, decades of investigation of the stress-timing/syllable-timing hypothesis have failed to reveal the necessary presence of corresponding durational regularities in the speech signal (see Cutler, 1991, for a review). The characteristic rhythm of a language is undoubtedly real, given that it plays a role in preferred poetic meters and, as we have seen, in lexical segmentation. However, its role is not to provide direct signals of word boundary location, but rather to establish a framework within which listeners can orient their hypotheses as to most probable word-boundary locations.

Prosody and lexical processing

In this section we consider the role of prosody in identifying words, and particularly in activating entries in the mental lexicon. One of the ways in which the phonological form of one word may differ from the form of another word is in prosodic structure, and thus prosodic information may serve to define a unique access representation in word recognition.

Because psycholinguistic models of spoken-word recognition have in recent years become computationally quite explicit, researchers concerned with the role of prosody in this process have given considerable attention to exactly how prosodic effects could be implemented. One major concern has been the domain of prosodic effects, and another the availability of the acoustic information which signals prosodic contrasts.

The issue of domain can be illustrated by comparing lexical tone as in Thai or Cantonese, and lexical pitch-accent patterns as in Japanese. Both are realized by variation in fundamental frequency (FO). But while pitch accent contrasts have a multisyllabic domain, tone contrasts may be realized within a single syllable. Thus the Cantonese CV sequence [si] with the high level tone 1 means *poem,* with the high rising tone 2 means *history,* with a low level tone 6 means *time* and so on. Tone can be realized on a single syllable; a syllable can be spoken in isolation with its tone unambiguously expressed. A tonal contrast

can thus be exemplified by comparing monosyllables. A Japanese pitch-accent pattern, on the other hand, essentially involves more than one syllable (or precisely speaking, more than one mora): an initial H (high) is always followed by a L (low), and vice versa. Thus Japanese pitch-accent patterns can only be exemplified by contrasting polysyllabic sequences.

Of course this description is in practice over-simplified, for both tone and pitch accent. Complicating factors include the fact that in Japanese, monomoraic words such as *hi* ("fire" in one realization, "day" in another) have the potential for contrasting realization in context even if this potential is neutralized when the words are spoken in isolation (Vance, 1987). With tone realization, on the other hand, there are context effects such as tone sandhi (Speer, Shih, & Slowiaczek, 1989, and Yuen, 1995, provide evidence that context affects tone identification in Mandarin and Cantonese respectively). Thus in effect a tone can sometimes only be perceived with reference to other syllables. Furthermore, in some tone languages (e.g., Mandarin) certain syllables can be unmarked for tone, with their realization dependent upon adjacent syllables; in others, such as some African languages (e.g., Sesotho), the picture is complicated by the fact that monosyllabic words do not occur. Nevertheless, the possibility of a contrast being realized in a multisyllabic domain raises questions for spoken-word processing, such as whether prelexical processing includes an explicit operation of comparison between syllables to extract pitch-accent information, and whether the time-course of arrival of the relevant information allows it to play a role in the earliest stages of lexical activation. Given that for both tone and pitch accent the relevant acoustic information is the FO pattern, evidence from studies of tone languages can help us understand the role of pitch accent in word recognition, and vice versa.

Related to this issue, clearly, is the question of just how the acoustic information is processed. Listeners make use of relevant acoustic information in spoken-word recognition as soon as it becomes available; for instance, they can efficiently use coarticulatory information in one segment to speed processing of the next (see, e.g., Whalen, 1991). Thus whenever it is proposed that the prosodic structure of words can constrain initial lexical activation, it is important to know exactly what acoustic information must be exploited for this to happen, and how such acoustic information is processed by the listener. If suprasegmental properties of the signal are involved then relevant information can be obtained from studies of processing of the same suprasegmental properties at other levels of the recognition process; if segmental information is involved, then relevant information can be obtained from studies of segment perception. Computationally explicit models of lexical activation demand that the time course of arrival of the relevant information be ascertained as accurately as possible.

Although these issues have been illustrated with examples from other languages, it is in English, not surprisingly, that most research on lexical prosody has been carried out; thus the prosodic structure involved has been lexical stress. More recently, the English work has been supplemented by analogous studies in Dutch, also a language with lexical stress. Stress is a good example of a phenomenon which can only be contrasted in a multisyllabic domain—it makes little sense to talk of the "stress pattern" of an English monosyllabic word for instance. As the following section will show, all of the questions outlined above have arisen in the research on stress.

Lexical stress. In English and Dutch, as indeed in most of the world's lexical-stress languages,

pairs of unrelated words differing only in stress pattern are extremely rare; thus although stress oppositions between English words of differing form class derived from the same stem *(decrease, perfect, conduct)* are common, there are very few such pairs which are lexically clearly distinct (such *as forbear,* or *insight / incite).* In principle, stress alone could serve to distinguish words, but in practice it rarely does.

Due to the greater acoustic reliability of stressed syllables, stress can affect recognition: stressed syllables are more readily identified than unstressed syllables when cut out of their original context (Lieberman, 1963), and distortions of the speech signal are more likely to be detected in stressed than in unstressed syllables (Cole, Jakimik, & Cooper, 1978; Cole & Jakimik, 1980; Browman, 1978; Bond & Games, 1980). When acoustic differences between stressed and unstressed syllables are relatively large, as they are, for instance, in spontaneous speech, detection of word-initial target phonemes is also faster on lexically stressed than unstressed syllables (Mehta & Cutler, 1988), but such differences do not arise with laboratory-read materials. Similarly, in a gating experiment (in which listeners hear words in successively larger fragments), stressed syllables are recognized earlier than unstressed syllables when the materials had been spontaneously spoken, but not when they had been read (McAllister, 1991).

This does not imply that stress differences per se are intrinsically salient to all listeners. Dupoux, Pallier, Sebastian, & Mehler (1997) found that speakers of French (a language which does not distinguish words by stress) had great difficulty processing stress contrasts in nonsense materials (recorded by a speaker of a stress language, Dutch); in a simple discrimination task they made many errors in deciding, for instance, whether a token *bopeLO* should be matched with an earlier token of *bopeLO* or *boPElo.* The same contrasts presented no difficulty to speakers of Spanish, a language which does distinguish between words via stress variation.

Studies of English vocabulary structure show that stress-pattern information could be of use in word recognition, despite the rarity of minimal stress pairs such *as forbear.* A partially-specified phonetic transcription (i.e., one which distinguishes only such phoneme classes as stop, fricative, vowel, etc.) will apply to a smaller candidate set of words if it includes stress-pattern information than if it does not (Aull, 1984; Huttenlocher & Zue, 1983; Waibel, 1988); but see Carter (1987) and Altmann and Carter (1989) for evidence that the set reduction due to stress information per se is, in information-theoretic terms, comparatively small. An automatic recognition algorithm operating on partially-specified phonetic representations performs significantly better with stress-pattern information than without (Port, Reilly, & Maki, 1988). In Dutch, studies of a 70,000 word corpus by van Heuven & Hagman (1988) found that words could on average be identified after 80% of their phonemes (counting from word onset) had been considered; when stress information was included, however, a forward search was successful on average given only 66% of the phonemes.

Nevertheless, stress information does not facilitate word recognition by English listeners: neither visual nor auditory lexical decision is facilitated by prior specification of stress pattern, nor does whether or not a bisyllabic word conforms to a canonical English word-class pattern (e.g., initial stress for nouns, final stress for verbs) affect how rapidly its grammatical category is judged (Cutler & Clifton, 1984). Note that listeners are certainly able to use knowledge of the canonical patterning in making "off-line" decisions, that is, responses that are not made under time pressure. A series of studies by Kelly (Cassidy & Kelly,

1991; Kelly, 1988,1992; Kelly & Bock, 1988) in which listeners are asked to use nonwords in a sentence as if they were words has shown that bisyllabic nonwords are more likely to be taken for nouns if they have initial stress, but for verbs if they have final stress. Similarly, English-speakers asked to use a verb as a nonce-noun choose a verb with initial stress, while for a noun acting as a nonce-verb they choose a noun with final stress.

Likewise, studies in Dutch using the (off-line) gating task have found effects of stress on listeners' word choices. Using minimal stress pairs (e.g. *SERvisch-serVIES)* presented in a sentence context, Jongenburger and van Heuven (1995a; see also Jongenburger, 1996) found that listeners' word guesses only displayed correct stress judgments for the initial syllable of the target word once the whole of that initial syllable and part of the following vowel were available. This relatively late differentiation of stressed and unstressed syllables in a free choice task contrasts with the demonstrations by van Heuven (1988) and Jongenburger (1996) that listeners can correctly select between two Dutch words with a segmentally identical but stress differentiated initial syllable (e.g., *ORgel* and *orKEST,* or a minimal pair such as *SERvisch-serVIES)* when presented with only the first syllable. Thus listeners in both English and Dutch are able to make use of stress information in off-line tasks. (Note that it is not known whether English listeners would be as successful as Dutch listeners in the forced-choice situation.) For free word recognition (as in the gating task) it may nevertheless be difficult to use stress pattern to guide choices if information from only a single syllable is available.

Many studies have, one way or another, addressed the effect of *mis-stressing* on word recognition. Puns are unsuccessful if they require a stress shift (Lagerquist, 1980). Deliberately mis-stressed English words are responded to more slowly in recognition tasks than correctly stressed words (Bond & Small, 1983; Cutler & Clifton, 1984), and the same is true of Dutch words (van Heuven, 1985). English listeners presented with English spoken by Indian speakers, in which correctly-placed stress is realized primarily by durational contrast but pitch movement may vary independently of duration, tend to interpret syllables with pitch movement as stressed; effectively, then, they are presented with misplaced stress, and they may report words which conform to the perceived stress, even though these conflict with the segmental information (Bansal, 1966).

In the case of the English experimental studies, the mis-stressing which was used often involved both segmental and suprasegmental manipulations. As we described earlier, pairs of English words with stress-pattern opposition usually also differ vocalically. Thus *SUBject* and *subJECT,* for example, have quite different vowels in their first syllables — the stressed syllable has a full vowel, while the vowel in the unstressed syllable is schwa. Just as the vowel difference in *sup* and *sop* is lexically significant, so may observed effects of stress primarily reflect the lexical significance of different vowels resulting from stress differences. Indeed, Slowiaczek (1990) found that if vowel quality is not altered, mis-stressing has no significant effect on the identification of noise-masked words; and Bond and Small (1983) found that mis-stressed words with vowel changes were not restored to correct stress in shadowing (indicating that subjects perceived the mis-stressed form and may not at all have accessed the intended word). Mattys and Samuel (1997) conducted a "migration" experiment, in which phantom word recognitions are induced by a combination of material presented separately to the two ears; they found that mispronunciation of a stressed syllable interfered with construction of the phantom percept, but mispronunciation of an unstressed syllable (whereby a reduced syllable became unreduced) did not.

To English listeners, the vowel quality distinction seems far more crucial than the prosodically defined distinction; cross-splicing vowels with different stress patterns produces unacceptable results only if vowel quality is changed (Fear et al., 1995). In Fear et al.'s study, listeners heard tokens of, say, *autumn,* which has primary stress on the initial vowel, and *audition,* which has an unstressed but unreduced vowel, with the initial vowels exchanged; they rated these tokens as insignificantly different from the original, unspliced, tokens. Bond's (1981) studies of "elliptic speech" —speech containing some systematic segmental distortion—showed that the manipulation which most inhibited word recognition was changing full vowels to reduced and vice versa. A reason why English listeners should closely attend to vowel-quality distinctions was suggested by Altmann and Carter's (1989) computation that the information value conveyed by phonetic segments in English is highest for vowels in stressed syllables.

To investigate *prosodic* effects on word recognition in a lexical-stress language, it is necessary to control for vowel quality. As long as both segmental and suprasegmental cues are available for listeners to distinguish stressed from unstressed syllables, it is impossible to rule out the argument that only the segmental cue is operative, in just the same way as *sup* is segmentally distinguished from *sop.* Although most unstressed syllables in English have a neutral (schwa) vowel, a reasonably large class of polysyllabic words with exclusively full vowels does exist. *Nutmeg* and *canteen* are two such words. In their mispronunciation experiment, Cutler and Clifton (1984) explicitly compared bisyllabic words in which the unstressed syllable contained schwa *(wallet, saloon)* with words like *nutmeg* and *canteen.* Word recognition was clearly inhibited by mis-stressing for the former group. The words with full vowels, however, were only harder to recognize when mis-stressed if their citation-form pronunciation had initial stress. That is, *nutMEG* was much harder to recognize than *NUTmeg;* but *CANteen* was not significantly more difficult than *canTEEN.* (In English, words like *canteen* are susceptible to stress shift in response to the demands of sentence rhythm; such words may be encountered sufficiently often in initially-stressed form for this form perhaps to have achieved the lexical status of an optional pronunciation.) Similarly, Taft (1984), using a monitoring task, found that initially-stressed words produced slower responses when mis-stressed *(cacTUS),* but mis-stressing of initially-unstressed words *(SUSpense)* actually led to response times which were somewhat faster than those with the correctly stressed words. Finally, Slowiaczek (1990) also demonstrated increased repetition difficulty for pronunciations like *nutMEG.*

In Dutch, however, some results suggested that mis-stressing of initially-unstressed words *(Plloot* instead of *piLOOT)* was more harmful to correct recognition than mis-stressing of initially-stressed words *(viRUS* instead of *Virus;* van Heuven, 1985; van Leyden & van Heuven, 1996). Van Leyden and van Heuven conducted their (gating) experiment in a parallel English version, in which, in contrast to the results from English described above, they observed equivalent effects of mis-stressing on both initially-stressed and initially-unstressed disyllables. (However, some of their English initially-unstressed words had reduced first syllables—*cigar, guitar*—which vitiates the cross-linguistic comparison; they also tested trisyllabic words, but again not in comparable manipulations across the two languages.) Another difference which they observed between Dutch and English was that a bias towards initially-stressed error responses, which appeared in both languages, was greater in English; they ascribed this to a greater probability of word-initial stress in English than

in Dutch polysyllables in natural speech (Cutler & Carter, 1987). Koster and Cutler (1997) carried out another study of the effects of mispronunciation in Dutch, in which they examined the effects of segmental mispronunciation and mis-stressing separately; the asymmetry reported by van Heuven and colleagues was replicated only when segmental mispronunciation was involved. Koster and Cutler did find significant adverse effects of mis-stressing involving no segmental mispronunciation (e.g., *coBRA* for *CObra*, *FATsoen* for *fatSOEN)*, suggesting again that stress, however realized, plays a role in word recognition in Dutch.

The process of word recognition includes several subsidiary operations, however, as we pointed out above; thus there are at least two ways in which prosody could be relevant in recognition. These correspond to the commonly drawn distinction between lexical access and lexical retrieval. On the one hand, lexical prosody (i.e., stress marking) could be an essential part of the access code by which lexical entries are activated; on the other, it could be part of the phonological code listed for a word in the lexicon and consulted only once access to the entry has been achieved. The mis-stressing results, even with on-line tasks, do not distinguish between these two possibilities. If prosodic information is present in the prelexically-used access code, *nutMEG* or *Plloot* could be hard to recognize because the initial access attempt will encounter no match, or (as suggested by van Leyden and van Heuven [1996] for Dutch) will encounter inappropriate matches, and an existing entry will only be activated after the code has been recomputed. If prosody does not play a role in access, however, *nutMEG* could be hard to recognize because the complete phonological form in the accessed lexical entry proves not completely to match the input.

One study of the lexically determinative function of stress is that of Connine, Clifton, and Cutler (1987), who asked listeners to categorize an ambiguous consonant (varying along a continuum between [d] and [t]) in either *DIgress-TIgress* (in which *tigress* is a real word) or *diGRESS-tiGRESS* (in which *digress* is a real word); the responses showed effects of stress-determined lexical status, in that /t/ was reported more often for the initial-stress items, but /d/ more often for the final-stress items. Here listeners could, it is clear, use the stress information (both that in the signal and that in their stored representations of these words) to resolve ambiguity in a difficult perceptual situation. Similar effects of lexical status are found with ambiguous consonants in CVC syllables, both in word-initial (Ganong, 1980) and word-final position (McQueen, 1991). However, evidence that a given source of information is used in a phonetic categorization task does not constitute evidence that it is used prelexically; the listeners in Connine et al.'s study had ample opportunity to consult the phonological code listed in two lexical entries, and to use the prosodic information contained therein to motivate their phonetic categorization decision.

If prosody participates in lexical access in much the same way that segmental identity does—that is, as part of the prelexical access code—then minimal stress pairs such *asforbear* should generate distinct access codes, and be, in practice, not confusable. Experimental evidence, however, suggests that the access code does not draw on prosodic information in English. Using the cross-modal priming paradigm (Swinney, 1979), in which listeners hear a sentence and at some point during the sentence perform a visual lexical decision, Cutler (1986) showed that pairs like *forbear* are functionally homophonous: *both* stress patterns, *FORbear* and *forBEAR* facilitate recognition of words related to *each* of them (e.g., *ancestor*, *tolerate)*. In other words, listeners did not distinguish between these two word forms in initially

achieving access to the lexicon. The prosodic information, although available in the signal, failed to exercise a directive function in constraining the prelexical access code to exclude one of the two possibilities. L. Slowiaczek (personal communication) has similarly found priming for associates related to both phrase-stress and compound-stress readings of strings such as *green house.*

On the other hand, a cross-modal priming study in Dutch, planned as a direct replication of Cutler's (1986) experiment, failed to find any significant priming at all from initially-stressed members of stress pairs *(SERvisch),* and inconsistent results for finally-stressed tokens *(serVIES;* Jongenburger & van Heuven, 1995b; Jongenburger, 1996). The inconsistent results from gating studies in Dutch, reported above, already suggested that the situation in Dutch might not be exactly the same as in English. Indeed, recent results from Dutch suggest that mis-stressing a word can prevent lexical activation. In word-spotting experiments (Cutler & Norris, 1988), in which listeners respond when they detect a real word embedded within a nonsense input, embedded words are detected less rapidly when they occur within a string which itself could be continued to form a longer word; thus English *mess* is detected less rapidly in *doMES* than in *neMES,* presumably because *doMES* could be continued to form the word *domestic,* while *neMES* cannot be continued to form a longer real word (McQueen et a1, 1994). Similarly Dutch *zee* (sea) is harder to spot in *muZEE* (which can be continued to form *museum)* than in *luZEE;* but only if *muZEE* is, like *museum,* stressed on the second syllable. If listeners instead hear *MUzee* and *LUzee* (i.e., the same strings but stressed on the initial syllable), then there is no longer a significant difference between these in detection time *for zee,* suggesting that there was in this case no competition from *museum* because it simply was not activated by input lacking the correct stress pattern (Donselaar, Koster, & Cutler, in preparation. Note that this experiment could not be replicated in English since English does not contain sufficient words beginning with two strong syllables and containing a single embedded word).

Donselaar et al.'s result suggests that in Dutch, at least, there may be on-line directive use of stress information in lexical access. The same conclusion can be drawn from the finding that the fragment *aLI-* will prime *aLInea* but not *Alibi,* and the fragment *Ali-* will prime *Alibi* but not *aLInea* (Donselaar, Koster, & Cutler, in preparation); this result was also observed with similar fragments of Spanish words (e.g., the first two syllables of *ARtico* or *arTIcolo)* presented to Spanish listeners (Soto, Sebastian-Galles, & Cutler, in preparation). These results from other languages — indeed, even the fact that other languages, unlike English, allow such experiments to be constructed—suggest that the failure to find similar evidence in English may arise from the peculiar redundancy of stress cues in English; stress information can nearly always be derived from segmental structure, and words can virtually always be distinguished by segmental analysis without recourse to stress.

This entire section, it should finally be noted, has dealt with free-stress languages. To our knowledge, no studies of the effects of mis-stressing have been carried out in a fixed-stress language such as Finnish or Polish. Indirect evidence is available from a word-spotting experiment in Finnish by Suomi, McQueen, and Cutler (1997), in which bisyllabic words (e.g., *palo)* were attached to preceding or following CV contexts *(kupalo, paloku);* all of the resulting trisyllabic nonsense items were spoken with the unmarked prosodic pattern for trisyllabic words, traditionally described as stress on the initial syllable. In a control experiment, the words were excised from their contexts and presented to listeners in a

lexical-decision task; for the words from which preceding contexts had been removed, this resulted in loss of the syllable which had nominally been stressed, and these words might therefore have been expected to be prosodically abnormal compared to those from following contexts. Listeners' responses showed no effect attributable to abnormality of this kind however; if anything, words like *palo* from *kupalo* were recognized slightly faster than words like *palo* from *paloku*. Suomi et al. suggested that the so-called initial stress of Finnish is actually a gradual drop in fundamental frequency and amplitude across the word, and that what is important for its correct realization is simply the relationship between consecutive syllables; this relationship would be unaffected by removal of preceding or following syllables. Fixed lexical prosodic structure may thus require no specifically prosodic perceptual processing at all in word recognition. However, the findings so far do not rule out the necessity of inter-syllable comparisons of prosodic realization in processing fixed-stress languages, and the effects of disruption of this expected pattern are as yet unknown.

Lexical tone. The results from experiments on English stress were interpreted in terms of the redundancy of cues to English stress: listeners can distinguish words by segmental analysis alone. But of course it is also possible that suprasegmental cues to word identity are for whatever reason inherently difficult to process, and that English listeners devote less processing attention to suprasegmental structure for this reason. An obvious source of relevant evidence is therefore the question of spoken-word recognition in languages in which words are distinguished via those types of information which cue English stress suprasegmentally—principally, duration and FO. Duration distinguishes words in languages with quantity distinctions (such as Estonian, which has a three-way vowel quantity distinction); however we know of no studies of the time-course of word activation and retrieval in such languages. FO distinguishes words in tone languages, especially Asian languages in which a highly restricted range of syllable structures is often combined with a relatively large tone repertoire. In such languages the information value of the FO contribution to the distinctiveness of the word form is high.

There is experimental evidence on how lexical tone information is processed in spoken-word recognition, but again it is not extensive by comparison with the literature on lexical stress. A categorization experiment by Fox and Unkefer (1985), using a continuum varying from one tone of Mandarin to another, confirms that listeners may of course use tonal information to determine word identity: the crossover point at which listeners switched from reporting one tone to reporting the other shifted as a function of whether the CV syllable upon which the tone was realized formed a real word when combined only with one tone or only with the other tone (in comparison to control conditions in which both tones, or neither tone, formed a real word in combination with the CV). The lexical effect appeared only when the listeners were Mandarin speakers; English listeners showed no such shift, and on the control continua the two subject groups did not differ. Such an effect of word/nonword status also appears, as described above, when it is the stress pattern—which, as we saw, appears not to be used prelexically in English — which determines word versus nonword status *(Tigress* vs. *diGRESS;* Connine et al., 1987). Fox and Unkefer's finding can, like Connine et al.'s, be explained in terms of exploitation of lexically stored information; thus although it is clear evidence that tone information is used in word recognition, it does not shed light on the time course of the processing of such information (e.g., whether tone plays a role in initial activation of word candidates or only in selection between them).

Further evidence that tone information contributes to word identification comes from a finding by Ching (1985, 1988) that identification scores for lip-read Cantonese words improved greatly when FO information was provided in the form of high-pass filtered pulses synchronized with the talker's larynx frequency (there was however very little improvement when FO information was provided for lip-read English words). Lexical priming studies in Cantonese also suggest that the role of a syllable's tone in word recognition is analogous to the role of the vowel (Chen & Cutler, in press; Cutler & Chen, 1995); in an auditory lexical-decision task, overlap between a prime word and the target word in tone or in vowel exercised parallel effects.

Nevertheless, there is evidence from experiments on the processing of Chinese languages that the processing of tonal information may be more error-prone than the processing of segmental information. In a study by Tsang and Hoosain (1979), Cantonese subjects heard sentences presented at a fast rate and had to choose between two transcriptions of what they had heard; the transcriptions differed only in one character, representing a single difference of one syllable's tone, vowel, or tone + vowel. Accuracy was significantly greater for vowel differences than for tone differences, and tone + vowel differences were not significantly more accurately distinguished man vowel differences alone. Taft and Chen (1992) found that homophone judgments for written characters in Mandarin were made less rapidly and less accurately when the pronunciation of the two characters differed only in tone, as opposed to in vowel; the response-time difference (though not the accuracy difference) was replicated in a second experiment in Cantonese. Repp and Lin (1990) asked Mandarin listeners to categorize nonword CV syllables according to consonant, vowel, or tone; the tonal categorizations were made less rapidly than the segmental decisions. Cutler and Chen (1997) found that in an auditory lexical-decision task, Cantonese listeners were significantly more likely erroneously to accept a nonword as a real word when the only difference between the nonword and a real word was in the tonal value of the second syllable. Such an error was moreover particularly probable when the FO onset difference between the correct tone of the real word and the erroneous tone on the nonword was small, so that the tone distinction was, in effect, perceptually hard to make. Similarly, in a same-different judgment task, Cutler and Chen found that Cantonese listeners were slower and less accurate in their responses when the only difference between two syllables was in their tone, than when a segmental difference was present. Again, responses were particularly slow and error-prone when the FO onset difference between the two tones was small.

This last effect also appeared in a study of the perception of Thai tones by Burnham, Kirkwood, Luksaneeyanawin, and Pansottee (1992): the order of difficulty of tone pairs presented in a same-different judgment task to English-speaking listeners was determined by the nominal starting pitch of the tones. In a subsequent study, Burnham, Francis, Webster, Luksaneeyanawin, Attapaiboon, Lacerda, and Keller (1996) compared same-different discrimination of Thai tones and musical transformations of the same tones, by speakers of Thai, Cantonese, and English. They found that Thai and Cantonese listeners discriminated the speech and musical tones equally well, but English listeners discriminated the latter significantly better than the speech tones (although the FO information was the same in both stimulus types). This latter result, suggesting that nonsegmental speech contrasts not found in one's native language are difficult to discriminate, adds to the findings of Dupoux et al. (1997) that stress contrasts are hard for speakers of nonstress languages, and of Nishinuma

(1994; Nishinuma, Arai, & Arusawa, 1996) that Japanese pitch-accent contrasts are difficult for non-native speakers of Japanese. Again using the same-different judgment task, Lee, Vakoch, and Wurm (1996) also found that English listeners had difficulty discriminating Cantonese and Mandarin tones; these authors also found that speakers of the two tone languages were better at discriminating tone contrasts of their own language than of the other language (although they were still better than the English listeners).

From the phonetic literature we know something of how tone is identified, at least in Chinese languages. Lin and Repp (1989), for example, report that identification of Taiwanese tones is based almost solely on processing of FO (height and movement), although tone and syllable duration do co-vary in this language (as in Cantonese, Kong, 1987; and in Mandarin, Kratochvil, 1971). Gandour (1981, 1983) similarly claims that three dimensions of FO are involved in tone identification in Cantonese: FO contour, direction, and height. Shen and Lin (1991), however, report that Mandarin tones 2 and 3, both of which end in a rise, are distinguished by the timing of the FO turning point within the syllable. Thus it is clear that tone identification in Chinese languages normally involves processing of FO, and it is possible that the processing may involve more than one dimension. Ultimately, however, tones are primarily realized upon vowels, and therefore they cannot be processed until the vowel information is available. Tonal information conveyed on a vowel and the vowel information itself are unlikely to be processed fully independently; classification of vowels in CV syllables is slower if the pitch of the syllable varies than if it is held constant, and likewise classification of pitch is slower if the vowel on which it is realized varies than if it is constant (Miller, 1978; Repp & Lin, 1990; Lee & Nusbaum, 1993). Yet vowels themselves can be identified very early; in a CV sequence the transition from the consonant into the vowel is enough for listeners to achieve vowel identification (Strange, 1989). Thus in a CV syllable in a tone language the various dimensions of the syllable would seem to have a fixed order in which they can be processed — consonant, vowel, tone. That is, tone decisions are slower than segmental decisions because the information on which a tone decision is based is not available until after the information relevant to the segmental decisions. The effect of this is particularly apparent in speeded response tasks, due to the pressure to respond quickly: in some cases subjects have issued their response before the tonal information has effectively been processed.

Lexical pitch accent. Pitch accent, as it is realized in Japanese words, resembles stress in that it involves a polysyllabic domain. However, it also resembles lexical tone in that it is realized principally via FO variation, and that minimal pairs of words differing only in prosody are not vanishingly rare. Japanese contains many more minimal accent pairs than languages such as English and Dutch contain minimal stress pairs. However, only meagre evidence is so far available on the processing of pitch accent in Japanese word identification. Otake et al. (1993) found no effects of pitch accent in their segmentation study: thus the first CV of a word was perceived equally rapidly and accurately irrespective of whether the word had HLL (e.g., *kanoko)* or LHH (e.g., *kinori)* accent pattern. Walsh Dickey (1996) conducted a same-different judgment experiment in which Japanese listeners heard pairs of CVCV words or nonwords which were either the same, or differed either in pitch accent or in one of the four segments. "Different" judgments were significantly slower for pairs varying in pitch accent than for pairs which varied segmentally, irrespective of the position of the segmental difference. Thus even a difference in the final vowel (at which time the pitch-accent pattern should also

be unambiguous) led to significantly faster responses than the pitch-accent difference. This finding is consistent with the findings on Cantonese tone showing that FO information becomes available later than cues to vowel identity; it may also (as Walsh Dickey, 1996, suggests) reflect a necessity that Japanese pitch accent be integrated across a multisyllabic domain.

More recent results suggest the possibility of a lexically directive function for pitch accent. Cutler and Otake (1996) presented Japanese listeners with single syllables edited out of bisyllabic words differing in accent pattern (e.g. *ka* from *baka* HL vs. *galea* LH, or from *kage* HL vs. *kagi* LH), and asked them to judge, for each syllable, in which of two words it had originally been spoken (effectively, whether it had H or L accent, since the two choice words were matched for phonetic context adjacent to the *ka).* The listeners were able to perform this task with an accuracy well above chance level, and, most interestingly, their scores were significantly more accurate for initial (80% correct) than for final syllables (68%). This suggests that pitch-accent information is realized most clearly in just the position where it would be of most use for listeners in on-line spoken-word recognition, although it of course does not of itself constitute evidence that the information is used in accessing the lexicon.[1] However, subsequent experiments by Cutler and Otake (in preparation) do suggest that pitch-accent patterns constrain word activation. In a gating study, listeners were presented with successively larger fragments of words such as *nimotsu* HLL or *nimono* LHH; their incorrect guesses from about the end of the first vowel *(ni-)* overwhelmingly tended to be words with the same accent pattern as the actually spoken word. And in a repetition priming study minimal pitch-accent pairs such as *hashi* HL and *hashi* LH did not facilitate one another's recognition. Together these results suggest that Japanese pitch accent is exploited by listeners in word activation, although, as it is cued by FO, the relevant information is only reliably available once a good part of the vowel upon which it is realized has been heard.

Summary

The evidence from word recognition is not yet sufficient to provide a complete picture of the role that prosodic structure may play in this process; there are still enormous gaps in our knowledge. Even for free-stress languages such as English and Dutch, despite the large volume of work reviewed in our section on stress, there are obvious experiments which have not yet been done; for fixed stress and all other varieties of lexical prosody the field is still wide open.

As we have outlined, there has been some progress in understanding how suprasegmental cues to word identity are processed. Research on the processing of tone suggests that tonal information is processed relatively slowly in comparison to segmental information. It may on occasion even fail to contribute to lexical access and selection. Again, there is considerable room for further empirical work in this area; for example, we know of no studies of activation of/competition between candidate words in which the role of lexical tone has been assessed.

[1] The task is similar to the forced-choice decision tasks used by van Heuven (1988) and Jongenburger (1996), in which Dutch listeners could also accurately decide from which of two specified words a syllable had been excised, although they did not accurately decide on the syllable's stress when given the same syllables without specific word prompts.

Further, as pointed out above, relevant evidence from studies of word activation in quantity languages is as yet lacking.

But the accumulated evidence on lexical stress does suggest that it is exploited by listeners in a number of languages, and not exploited in English where vowel-quality information essentially provides all the cues to word identity which prosodic processing could provide. Thus the evidence is consistent with the hypothesis that prosodic contrasts are disregarded by listeners where they are fully redundant, and differ principally in the degree to which listeners may be able to get away with ignoring them. Presumably, prosodic contrasts do not differ from segmental contrasts in this respect.

At what point in the word-recognition process does stress play a role? The newest studies involving measures of activation and competition suggest that stress may have a role in the initial activation of lexical entries in those languages where it contributes significant information to word identification; but more experimental evidence is needed, from more languages, and explicit modeling of the contribution of stress in a computational model of spoken-word recognition is also overdue.

PROSODY IN THE COMPUTATION OF SYNTACTIC STRUCTURE

Introduction

Syntax and prosody are closely related. The suprasegmental characteristics of words may, for instance, be influenced by position in syntactic structure: Greater FO movements and longer segmental durations are observed before major syntactic boundaries (Vaissiere, 1974,1975; O'Shaughnessy, 1979; Cooper & Sorensen, 1981, for FO contour variations in French and English; Oiler, 1973; Klatt, 1975,1976; Cooper, 1976; Bouwhuis & de Rooij, 1977; Cooper & Paccia-Cooper, 1980; Garro & Parker, 1982; Kutik, Cooper, & Boyce, 1983, for segmental durations and pausing in English and Dutch). There is considerable similarity across languages in the prosodic correlates of utterance position (see Vaissiere, 1983, for a review of relevant literature). Listeners can make use of the syntax-prosody relationship: major syntactic boundaries may be accurately located from prosodic information alone, no lexical information being provided (Collier & 't Hart, 1975, using hummed versions of utterances; de Rooij, 1975, using spectrally scrambled speech; de Rooij, 1976, using reiterant speech; Collier, de Pijper, & Sanderman, 1993; de Pijper & Sanderman, 1994; Sanderman, 1996, using filtered speech). Lehiste and Wang (1977) and Kreiman (1982) used a corpus of spontaneous speech, suppressed the segmental information, and then asked listeners to locate sentence and paragraph boundaries. Even though their accuracy rate was rather low (perhaps because such spontaneous speech involved some interruptions and incomplete utterances), there was strong agreement across the listeners'judgments, which suggests shared knowledge about how such boundaries are acoustically marked in speech.

However, syntactic and prosodic structures are not isomorphic (for discussion, see Steedman, 1990; Hirst, 1993; Inkelas & Zee, 1990; for experimental evidence, Gee & Grosjean, 1983; Monnin & Grosjean, 1993; Ferreira, 1993; also Shattuck-Hufnagel & Turk, 1996, for a review). The prosodic structure of an utterance can indeed be seen as a grammatical entity in its own right, requiring its own parsing and, importantly, allowing its own ambiguities (Beckman, 1996). The question addressed in the next section of this review is whether the

perceptual grouping of words into higher-order units defined by prosodic structure helps syntactic analysis, even without prosody-syntax isomorphism and unambiguous mapping.

Determination of higher-level constituent structure

The question whether prosody participates in the on-line segmentation of speech into constituents was addressed in the very earliest days of modern Psycholinguistics. Research with a number of different paradigms showed evidence of an active syntactic structuring undertaken by the listener. For instance, in several studies which will be described in more detail below, sentences were presented to listeners with, at some specific point, the presentation disrupted by (a) a click, either superimposed on the speech or presented to the ear other than the one the speech is presented to, or (b) a switch of the signal from one ear to the other. The listeners' task was to indicate when the click was presented or when the switch occurred, either by marking a written text at the end of the sentence presentation, or by writing the sentence down and indicating on this transcript where the click or switch occurred. The main assumption of this paradigm is that a perceptual unit tends to preserve its integrity by resisting interruptions. A segmentation process is evidenced by a reliable tendency for the subjective location of the interfering stimulus to migrate toward the boundary of the segmented unit.

Just such evidence of grouping into constituent-structure units was found: clicks tended to migrate towards major syntactic breaks. The authors involved in these studies at first attempted to rule out potential use of prosodic cues in their quest to observe syntactic processing alone; thus Abrams and Bever (1969) presented utterances with what they referred to as "subdued" intonation while Fodor and Bever (1965) and Garrett (1964) controlled for the presence of pauses. Later studies explicitly attempted to evaluate the respective influence of prosody and syntax in click migration. Thus Garrett, Bever, and Fodor (1965), and Bever, Lackner, and Kirk (1969) constructed stimuli in which prosody and syntax conflicted; they recorded pairs of sentences sharing a lexical common part but with different syntactic structure inducing prosodic breaks at different locations (e.g., *[In her/Your] hope of marrying Anna was surely impractical).* They then cross-spliced the common part (here: from *hope* onwards) from one sentence of the pair to the other; for the cross-spliced versions, the groupings induced by the sequence of words differed from those induced by prosodic cues. The intact and cross-spliced versions were presented to listeners, who had to judge the location of a click. A similar click-placement pattern was observed for both spliced and unspliced versions of the sentences: the clicks were reported to have occurred at boundaries consistent with the syntactic structure. The authors concluded that a click-localization pattern was determined by the syntactic structure of the sentence, and not by the prosodic-boundary information. This suggests that prosodic information is not powerful enough to rule out syntactic grouping.

Wingfield and Klein (1971) (see also Wingfield, 1975a, 1975b; Wingfield, Buttet, & Sandoval, 1979) attempted to ascertain whether prosody made any contribution at all to the perception of constituent structure. Using the dichotic switch methodology, again with spliced stimuli of the type used by Bever et al. (1969), they observed that when prosody and syntax were in conflict, the switch point reports migrated neither towards the prosodic break nor, in contrast to the findings of Garrett et al. and Bever et al., towards the major syntactic break. Wingfield (1975a) proposed that perceptual segmentation is determined

primarily by syntactic structure, but with accompanying acoustic patterns serving to mark the underlying structure or aid in syntactic resolution. Perceptual segmentation is thus seen as a stage of processing in which prosodic cues directly aid syntactic analysis. If prosody and syntax conflict, the perceptual grouping seems to be affected by the discrepancy.

Geers (1978) provided further support for this argument, in a study using the click-localization paradigm. She recorded sentences where the prosodic break was naturally realized by a speaker either at the major syntactic boundary (a clause boundary, indicated by *II)* or at a minor syntactic boundary (a phrase boundary, indicated by /) (e.g., *Because it rained II the picnic I will be canceled),* and matched these with "nonsyntactic" utterances, consisting of strings of letters of the alphabet, containing prosodic breaks at the same positions as in the sentences, defined in terms of number of syllables. She found that both types of prosodic boundary attracted clicks in the sentences, but click reports in the nonsyntactic stimuli were not affected by prosodic-boundary placement. This finding suggests that syntactic cues are essential for the parsing of speech. When an intonational boundary and the major syntactic boundary coincide, there is a tendency to locate interrupting stimuli in the clause boundary, which is also marked by a prosodic break; prosody reinforces the perceptual effect rather than directs it.

This series of studies suggests a supporting, rather than a leading, role for prosody in the grouping of words into constituents. However, click localization is a task which relies on listeners' memory for where the extraneous signal of the source switch occurred in the speech signal. The responses, that is, are not made on line; thus the results do not directly address the issue of how each word is incrementally attached to the preceding context, and when and how syntactic and prosodic information are respectively used. It is possible that prosodic information is indeed exploited at an early processing stage, but that when syntax and prosody conflict, the prosodic information is overridden (at the later stage tapped by these memory tasks) by syntactic processing. Nor does the click-location task provide a measure of syntactic processing; at most, it indicates groupings resulting from processing of whatever kind may have intervened. Some of the more recent work on prosodic cues to syntax, as described later in the sub-section on the resolution of local ambiguity, has employed tasks which attempt to tap early stages of syntactic processing and to ascertain whether prosody can play a role at that point.

The resolution of global ambiguity

I read about the repayment with interest—does the prepositional phrase *with interest* refer to my reading or to the repayment? *La petite brise la glace*—NP V NP "The little girl breaks the ice," or NP Pro V "The little breeze freezes her"? Such sentences are globally ambiguous in that as a whole they admit of more than one interpretation; this type of ambiguity (also called, by Beach [1991] "standing," as opposed to "temporary," ambiguity) is not resolved by the occurrence of further linguistic information within the sentence. A number of studies have analyzed whether multiple interpretations of a spoken utterance can be disambiguated by prosodic information.

Listeners can in many cases correctly determine which interpretation of the sentence was intended by the speaker (Lehiste, 1972,1973; Wales & Toner, 1979; Price et al., 1991; Ferreira et al., 1996), by exploiting the prosodic correlates of syntactic breaks referred to in the *Introduction* to this section. Such use of prosodic boundary cues can override strong

preferences for one reading rather than another (Misono, Mazuka, Kondo, & Kiritani, 1997; this study, in Japanese, contrasted sentences which were for instance ambiguous between the assignment of a modifier "drunken" to a father or a baby). Avesani, Hirschberg, & Prieto (1995) showed that English, Spanish, and Italian speakers use intonational patterns to differentiate the interpretation of potentially ambiguous sentences, the strategies differing across languages. In Japanese, the absence of morphological marking on relative clauses leads to ambiguities between a simplex sentence and the subordinate clause of a complex NP. Venditti and Yamashita (1994) showed that the distinction between those two interpretations is expressed by lengthening, and lower amplitude and FO contour, on the final morae of the simplex sentence, compared to the complex construction. Those acoustic cues were salient enough for listeners to decide on the syntactic structure of the ambiguous string. In Korean, prosodic phrasing can help disambiguate "wh"-words that can signify either wh-pronouns (e.g., *who)* or indefinite pronouns (e.g., *anyone;* Jun & Oh, 1996).

Listeners' boundary decisions may be based on durational information — that is, preboundary lengthening (Lehiste, 1973; Lehiste, Olive, & Streeter; 1976; Scott, 1982; Warren, 1985, all for English; Nooteboom, Brokx, & de Rooij, 1978, for Dutch), as well as on pitch-contour variation, usually a preboundary fall-rise or rise (Cooper & Sorensen, 1977; Streeter, 1978; Beach, 1991; Wightman et al., 1992, for English; Bruce, Granstrom, Gustafon, & House, 1992, for Swedish). Amplitude has been shown to be a less reliable cue for syntactic decision-making than duration and pitch (Streeter, 1978, for English; Sorin, 1981, for French; but see Scholes, 1971, for evidence that listeners can use amplitude cues in English). These studies have in general considered syntactic structure in more detail than the studies discussed in the previous subsection, *Determination of Higher-Level Constituent Structure;* for instance, many studies have contrasted boundaries of differing strength as defined by position in the syntactic hierarchy (see Swerts, 1997, for a recent discussion of prosodic-boundary strength). Acoustic analyses by Price et al. (1991) and Wightman et al. (1992) showed that the amount of preboundary lengthening in English is related to the strength of the boundary, for relatively low and medium levels of boundary strength. For major prosodic boundaries (intonational-phrase boundaries), the strongest cues are conveyed by pitch rather than by duration.

Schafer (1995) used a question-answering task to establish which representation listeners had computed of an ambiguous sentence such as *Paula phoned her friend from Alabama,* produced with various alternative prosodic realizations of boundaries. Her results suggested an indirect relationship between the boundary realization and the interpretation; with either a late boundary (in this example, between *friend* and *from)* or no boundary at all, the PP tended to be interpreted as attached to the higher node and not to the noun phrase. Schafer suggested that prosodic phrases do not cue syntactic phrases directly, but their structure is checked for compatibility with a syntactic representation.

In fact, disambiguation of global ambiguity is not always possible; listeners' ability to resolve such ambiguity seems to vary depending on whether speakers were aware of the possibility of multiple interpretations when they produced the utterance (Lehiste, 1973; Wales & Toner, 1979; Allbritton, McKoon, & Ratcliff, 1996; Mazuka, Misono, Tadahisa, & Kiritani, 1997). Speakers produce few cues to some syntactic structures when the context is strongly biased towards one interpretation (Straub, 1996; Mazuka et al., 1997). Moreover, some types of ambiguity are consistently less accurately resolved than others. Read, Kraak,

and Boves (1980) found that Dutch verbs differed in the extent to which they allowed disambiguation of who-questions via shift in sentence accent; thus *wie zoent de vrouw?* ("who is [kissing the woman/the woman kissing] ") and *wie kijkt de vrouw aan?* ("who is [looking at the woman/the woman looking at] ") are both ambiguous, but placing an accent on the verb causes the former to be more readily accepted than the latter as questioning the object of the action. Lehiste (1973) found disambiguation of surface structure ambiguities *(/saw the man with the telescope; German teachers)* to be easier than disambiguation of deep structure ambiguities *(John doesn 't know how good meat tastes).* Ferreira et al. (1996) also found differences between structural types.

Nespor and Vogel (1983) proposed that two interpretations of the same syntactic structure can be distinguished only if their prosodic constituent structure differs. As research in prosodic phonology has shown (see Shattuck-Hufnagel & Turk, 1996 for a review). prosody marks boundaries and delimits constituents which do not always correspond to syntactic units. If the two interpretations of an utterance have the same syntactic constituency. the prosodic structure must also be identical, argued Nespor and Vogel, since the first determines the second. But in all other cases (i.e., when the constituent boundaries are differently located, irrespective of whether the labels of the constituents are identical or different), the prosodic phrasing can vary in accord with the syntactic constituency and labeling, and the two interpretations may be prosodically distinguished. Nespor and Vogel supported their argument with perceptual evidence from Italian. As predicted, disambiguation was easiest when the two interpretations differed in the location of intonational-phrase boundaries *(Quando Giorgio chiama suo fratello e sempre nervoso* "When George calls his brother [is always nervous/he is always nervous]"), and was also possible when they differed only with respect to phonological-phrase boundaries *(La vecchia legge la regola* NP-V-NP "The old lady reads the rule" versus NP-Pro-V "The old law regulates it," as in the French case at the beginning of this section); but when both intonational- and phonological-phrase structure were identical, even though the labels of syntactic constituents differed *(Andrea taglia la radice e Luca la pianta* "Andrew cuts the root and Luke [Det-N the plant/V-Pro plants it]"), the two interpretations could not be reliably distinguished.

These studies suggest that some syntactic information can be extracted from prosodic cues, in particular information about the location of syntactic-constituent boundaries. However, they do not address the issue of when and how this prosodic information is integrated with the linguistic content of the sentence. Nicol and Pickering (1993) approached this question, using an on-line task to study global ambiguities of the type: *The receptionist informed the doctor that the journalist had phoned about the events.* Here the embedded clause introduced by *that* may be construed either as a sentential complement or as a relative clause. Nicol and Pickering focused on the time-course of disambiguation (an issue which has mostly been raised in studies of local ambiguity, as will be seen in the following subsection). Their study investigated, via use of a cross-modal priming technique (see Nicol, Fodor, & Swinney, 1994), whether both interpretations are simultaneously constructed or whether one interpretation is preferred at an early processing stage. If the relative-clause interpretation is available, the verb *phoned* should activate its object, namely *doctor:* evidence of this reactivation would be provided by faster lexical decisions to a word semantically related to *doctor* (PHYSICIAN) presented just after *phoned.* If the sentential-complement interpretation is activated, no such priming effect should be round. They

presented two versions of the sentence, recorded in each of the two interpretations, thus with two potentially differing prosodic patterns. The results showed facilitation (compared to a baseline condition) for the related word given the sentence read with the relative-clause interpretation, but no significant facilitation given the sentential-complement interpretation. This difference indeed suggests a possible role of intonation in favoring one interpretation at *phoned* (the word which in fact bore most of the prosodic differences between the two readings), though it does not entail that prosodic information forces one analysis rather than another at the point of ambiguity.

The resolution of local ambiguity

In fact, it is likely that relatively few of the utterances we hear are globally syntactically ambiguous in this way. Nevertheless, a great many sentences will contain temporary, or local, ambiguity with respect to how lexical units are attached or related to the preceding ones. Resolution of local ambiguity offers scope for integration of prosodic with syntactic information on-line, as the sentence unfolds, and by virtue of being widespread it could represent a significant opportunity for the exploitation of prosodic information in sentence processing. This issue has been the subject of much recent research. Consider such sentences as *John believes Mary implicitly; John believes Mary to be a professor*. Neither is ambiguous but both sentences begin with the same three words, so that within that initial three-word string in each sentence there can be said to be local ambiguity: the relationship between *believes* and *Mary* depends on what follows. Numerous recent studies have addressed the question of whether prosodic information can be used on-line to resolve such local, or temporary, ambiguities. In such studies listeners have typically been presented with the portion of a sentence which preceded disambiguating lexical information, and an attempt has been made to ascertain which syntactic analysis was available to them.

In some cases the structural ambiguity involved has been relatively simple. For instance, Grosjean (1983) examined the role of prosodic information as a predictor of utterance length, defined as number of prepositional phrases attached to a simple verb phrase. Sentences of various lengths were used, for example, *Earlier my sister took a dip / in the pool / at the club / on the hill.* Grosjean presented listeners with the potentially final word (e.g., *dip),* plus its entire preceding context, and found that they could successfully differentiate between the options that the sentence contained no more words, three more words, or six more words. (The option that nine more words were to follow was not reliably distinguished from six more words.) A key-press task, in which subjects were asked to listen to the sentence fragments and indicate when they believed the sentence would have ended, showed similarly that subjects could reliably distinguish the same three categories. Acoustic measures of the potentially final words (especially relative FO and syllable duration) seemed to correlate well with the results. In a follow-up study, Grosjean and Hirt (1996) tested and confirmed the implication of these correlations, namely that the subjects in Grosjean's (1983) study had made use only of information realized in the potentially final word. They also found the use of this prosodic information to be language-specific since French listeners, unlike English listeners, could not differentiate between English sentences that continued with differing numbers of added phrases, although they could tell whether a sentence ended or not. As English listeners were only found to make reliable predictions from the potentially final word, Grosjean and Hirt suggested that when syntax or semantics

informs listeners that the sentence is continuing, they may make less use of prosodic information concerning sentence length.

Further studies have explored the role of prosody in choosing at an early stage of processing between alternative likely syntactic structures for a partial utterance. To tap into early processing, these studies have used response-time methods. Studies of written language processing (see e.g., Mitchell, 1994, for a review) have suggested that some syntactic analyses are likely to be preferred over others. In spoken-language processing, however, such preferences may be overridden, since prosodic information will be available on line and will effectively disambiguate the partial string. A study by Marslen-Wilson, Tyler, Warren, Grenier, and Lee (1992) focused on the same case of attachment ambiguity illustrated at the beginning of this section, in which a verb can take either a direct object complement or a clause complement *(The workers considered the last offer from the management [of the factory /was a real insult]).* At issue is Frazier's (1978) Minimal Attachment proposal, according to which the preferred parse of such sentences attaches the NP *the last offer* as direct object to the sentence structure already in construction rather than beginning a clausal complement, that is, a new, embedded, sentence. In the Marslen-Wilson et al. (1992) study, the entire overlapping portion of the sentences (in the example, until *management)* was presented auditorily, followed by a visual probe, which was either an appropriate or an inappropriate continuation for the auditory sentence. Subjects had to read out (or "name") the visual probe as quickly as possible. RTs to the probe (which was always appropriate for the clause-complement interpretation, i.e., *was* in the example presented here) were faster in the clause prosody condition than in the direct-object prosody condition. The authors concluded that prosodic information is used at an early processing stage to resolve potential ambiguities in the structural interpretation of the utterance, favoring one syntactic interpretation over the other. Moreover, as the clause-complement interpretation was the one predicted to be disfavored by Minimal Attachment, Marslen-Wilson et al. claimed that prosodic cues to structure could override the operation of such preferences.

Watt and Murray (1996), however, questioned this conclusion, and also pointed to the partial nature of the experimental design used by Marslen-Wilson et al.: only one sentence version, instead of both of the two versions which are supposed to provide different prosodic cues, and only one visual probe, instead of appropriate versus inappropriate probes. Watt and Murray further queried the use by Marslen-Wilson et al. (1992) of an additional task in which listeners had to rate the goodness of each visual probe as a continuation of the spoken sentence; this, they argued, might have led subjects to process the sentence and the probe in an unnatural way. Watt and Murray therefore conducted the same experiment without this additional task but with the missing experimental conditions: two prosodic versions of each ambiguous string, and two probe words. Their results did not indicate any effect of the prosodic sentence versions on naming latencies. Nor did any effect appear when the experiment was replicated using a cross-modal lexical-decision task (in which lexical decisions to the probe were expected to be faster when the probe is consistent with preceding linguistic material). Watt and Murray concluded that the prosody of the ambiguous portion of the sentence did not constrain syntactic interpretation.

Warren et al. (1995) investigated the use of prosodic information in the resolution of closure ambiguities. Listeners were presented with the beginnings of sentences such as *Whenever parliament discusses Hong Kong problems,* for which *problems* can belong to

the object NP of *discusses*, or can be the subject of a following clause *(e.g., problems are solved instantly)*. The prosodic information they could use was two-fold: (1) presence or absence of a boundary tone at the end *of Hong Kong* or *of problems*, the presence of such a tone being a possible cue for a clause boundary; (2) the stress placement in stress-shift items such as *Hong Kong*.[2] Stress placement for these items is linked to the syntactic and prosodic structure of the utterance: A stress shift (i.e., *HONG Kong)* would indicate the presence of a following word in the clause *(problems)*, and hence a late closure. Although stress shift is not obligatory, and the absence of stress shift can be acceptable even in the case of a late closure, this information has been shown to be used by listeners (Grabe & Warren, 1995). In Warren et al. (1995), the stress-pattern and prosodic-boundary information were either consistent or conflicting. The study, like that of Marslen-Wilson et al. (1992) used the cross-modal naming paradigm; the probe presented for naming was always appropriate for the early-closure reading (here, *was)*. Warren et al. found that the naming latencies were slower when the prosodic boundary cued a late-closure reading. The effect of stress shift was less clear; it varied across sentences, and was for instance not used as a syntactic cue when there was also a possible reading in which the stress-shifted word was contrastively accented.

The same criticism raised by Watt and Murray (1996) of Marslen-Wilson et al.'s (1992) study could be made here: To test whether prosody determines the syntactic commitment made at the ambiguous point, one needs to show that prosody can induce both late and early closure, and can act to render a probe either appropriate or inappropriate. Marslen-Wilson et al.'s (1992) and Warren et al.'s (1995) studies certainly established that the processing of spoken ambiguous sentences does not systematically follow the principles established by Frazier and her colleagues (Frazier, 1978; Frazier & Fodor, 1978), and the contribution of prosodic information is a plausible source of this inter-modality difference. However, such principles had been already questioned in written-language studies (e.g., Kennedy, Murray, Jennings, & Reid, 1989); and given that Marslen-Wilson et al.'s (1992) findings were not replicated by Watt and Murray (1996), it seems impossible to draw firm conclusions regarding the role of prosody in syntactic parsing from this series of studies.

Note that different conceptions of what might be the role of prosody in syntactic parsing are addressed in those studies. The claim for a complete design, like the one that Watt and Murray (1996) used, rests on the assumption that prosody is a linguistic structure conveying information to the parser on its own. Hence, what is tested is whether prosody can force one parsing or the other, depending on the acoustic pattern. Another conception of prosody's role is the provision of information that the parser can use, although this information can be ambiguous. The simple fact that the results differ from what has been found with written language provides evidence that prosody is used in parsing spoken sentences, but it does not necessarily entail that prosodic information forces a particular parse.

In a further study, also using the cross-modal naming task, Grabe, Warren, and Nolan (1995) investigated local ambiguity involving adverbial versus clausal coordination *(Rachel smiled politely but [coldly/Kelly wasn't fooled])*. Naming of visual probes consistent with

[2] Stress shift refers to a reversal of the prominence pattern in a word with more than one full vowel and the strongest prominence in citation form on the last syllable, when it is followed by a word beginning with a strong syllable; see Liberman & Prince, 1977; Gussenhoven, 1991.

a clausal coordinate (KELLY) was elicited immediately at the end of the ambiguous portion of the sentences *(...but);* responses were faster when the ambiguous portion had been realized with the prosodic structure preferred for the clausal coordinate than when it had been realized with the prosodic structure preferred for the adverbial coordinate. Interestingly, Grabe et al. realized two versions of prosodic structure appropriate for each of the possible structures; in separate rating tasks listeners judged one version as preferable for each structure, while the other version was judged to be "marked" for a special context (e.g., such a structure would be acceptable in the case of, say, contrastive accent or topic change). The "marked" prosodic structure for the clausal coordinate (consistent with topic change) did not facilitate naming of the probe consistent with the clausal structure. In this study, therefore, listeners were sensitive not only to the syntactic structures consistent with the prosodic information, but also the relation of these syntactic structures to a wider context.

A somewhat different approach to local ambiguity was taken in a set of studies examining listeners' explicit judgments about the syntactic structure of potentially ambiguous fragments. Instead of using response-time methods to tap into processing at a particular point in the signal, the signal is effectively stopped before the point at which it becomes unambiguous. Beach (1991) presented listeners with the initial part of ambiguous sentences, again of the type used by Marslen-Wilson et al. (1992) and Watt and Murray (1996), where the ambiguity involved direct-object versus clause-complement attachment: *Mary suspected her boyfriend [immediately/was lying].* These initial fragments were either short *(Mary suspected),* or longer *(Mary suspected her boyfriend).* Listeners were asked to decide from which complete sentence the fragment had been excised. Beach found significant differences between the judgments, with the percentage of direct-object judgments being higher when the sentence had been produced with this interpretation than with the sentence-complement interpretation; she concluded that prosody can disambiguate very early in the sentence, before lexically disambiguating information is available. Although the difference between the two percentages was significant, the listeners' accuracy was in fact very low: direct-object interpretations of short fragments were chosen in 56.5% of cases for the direct-object prosody, compared to 42% for sentence-complement prosody, with the corresponding means for long fragments being 39% and 54%. A replication of Beach's experiment by Stirling and Wales (1996) found even smaller effects of prosodic information on judgments for short fragments (56.5% vs. 53%), and no difference at all for long fragments (50.1% for both). The absence of any distinction between the two readings for long but still ambiguous fragments must call into question the claim that prosodic cues have a cumulatively integrative effect on the listener's ability to interpret syntactic structure. It is regrettable that very few studies in this area have made any attempt to establish whether different productions of an ambiguous sequence do indeed contain different prosodic patterns. Clearly, listeners need an actual difference in prosodic realization if they are to distinguish local ambiguities via prosody; examination of the prosodic characteristics of each version of such a structure is thus crucial. Only if the two versions can be shown to provide contrastive prosody can one pose the question whether this information is used by the parser to decide syntactic attachment.

Some local-ambiguity studies have involved explicit manipulation of prosodic patterns. Thus Speer, Kjelgaard, and Dobroth (1996) hypothesized that the presence of a prosodic boundary at a potential syntactic-boundary location leads the parser to close the current syntactic constituent, leading to resolution of early/late closure ambiguity (*Whenever the*

guard checks the door [is/it's] locked). They constructed versions of such sentences with cooperative prosody (i.e., pitch excursion, lengthening, and silence at one or other syntactic boundary) and with conflicting prosody (by cross-splicing the two versions). In addition, there was a baseline version of each sentence, in which the region of the temporary syntactic ambiguity contained no prosodic boundary and for which pretests established that region as equally acceptable for both readings. (Note that the existence of a "neutral" prosodic version, i.e., a prosody equally acceptable for both syntactic structures, indicates that the parsing of prosodic structure can be ambiguous; Beckman, 1996). With these sentences listeners performed a simple comprehension task (press a button as soon as they have understood the sentence), as well as a cross-modal naming task. We will discuss the results for the latter, although both tasks led to equivalent conclusions. For the baseline condition, there was an advantage for the late-closure interpretation, consistent with results from reading experiments, and explained in terms of the Late Closure principle (Frazier & Rayner, 1987) or in terms of lexically-associated frequency information favoring the verb's transitive interpretation (MacDonald, Pearlmutter, & Seidenberg, 1994). For the two other conditions, the naming latencies in general showed a clear facilitation (compared to the baseline) when probes fitted the prosodic-boundary information (cooperative condition), and a clear interference effect when probes fitted the syntax but not the prosodic-boundary information (conflicting condition). Thus, when a prosodic boundary occurs at a point of syntactic ambiguity, it can determine the choice of parse. However, and importantly, there was no significant difference between the cooperative prosody condition and the baseline condition when the interpretation was late closure, implying that the availability of an appropriate prosodic boundary did not speed processing of the preferred interpretation.

In the last of their experiments, Speer et al. (1996) used less marked prosodic-boundary cues: A high phrase accent and no pause, which effectively created an intermediate-level prosodic boundary. They found the same overall result pattern, although the interference effect in the conflicting-prosody conditions, while still significant, was weaker. The influence of prosodic information on syntactic parsing therefore appears to be sensitive to the degree to which the prosody resembles a typical syntactic boundary.

Speer et al. (1996) hypothesized that the presence of a prosodic boundary at a potential syntactic-boundary location leads the parser to close the current syntactic constituent. Pynte and Prieur (1996) attempted to explain this process in the case of ambiguous attachment of a prepositional phrase (either attached to the verbal phrase or to the noun phrase: *The spies inform the guards of the [conspiracy/palace]).* They hypothesized that the presence of a prosodic break acts to block the minimal-attachment commitment of that phrase to the previous constituent, and tested this by presenting listeners with such sentences containing either one or two inserted prosodic breaks. The listeners detected a prespecified word target (which, in the experimental items, was always the last word of the sentence — *conspiracy I palace).* The findings did not fully support the hypothesis. NP attachments *(of the palace)* were facilitated by a break inserted after the verb, and VP attachments by an additional break before the PP, but a break before the verb did not inhibit NP attachment. This led Pynte and Prieur to propose instead that prosodic-break information is used to group words into low-level syntactic constituents rather than to place closure.

This view is shared by Kennedy, Murray, Jennings, and Reid (1989), and Murray, Watt, and Kennedy (submitted). According to their proposal, the parser would apply the strategy

of building low-level constituents and holding them unattached, until morphosyntactic disambiguating information becomes available. Such a "chunking" procedure would minimize memory load without committing the listener to one (perhaps incorrect) analysis. Prosodic-boundary information would play a major role in the "chunking" process. Schafer (1995) concluded from her study described above that prosodic boundaries do not directly indicate any kind of attachment, but are rather used as cues for the construction of a prosodic representation, for example, as marking the edge of a prosodic phrase. The incoming material would be incrementally incorporated into a well-formed prosodic representation, as part of a prosodic phrase, and structured within that phrase before it is attached to the larger syntactic representation. Current phonological theories of the syntax-prosody mapping allot a central role to the edges of syntactic constituents in defining prosodic domains (Selkirk, 1986; Rice, 1987). The psycholinguistic studies reviewed here strongly support this view. Note that the separate prosodic parse proposed by Beckman (1996) here is accorded temporal priority over the syntactic parse in processing; this issue is one where further investigation seems called for.

Finally, Schafer, Carter, Clifton, and Frazier (1996) investigated the role of focal accent in determining syntactic attachment. In sentences such as *The detective eyed the entrance of the house that shows clear signs of damage*, the relative clause could in principle be attached to either *entrance* or *house*. Schafer et al. found that syntactic parsing is influenced by focal accent: listeners' responses to questions such as *What showed signs of damage?* showed that they preferred to attach the relative clause to whichever noun was accented. As the following section will describe, accent conveys informational focus, and listeners actively seek focussed words; Schafer et al.'s finding suggests that the most important information is considered the most likely to be elaborated, and that syntactic attachment may therefore be consequent upon decisions about semantic structure. Certainly it is consistent with a notion of prosody as a structure closely related to both syntax and information structure (see, e.g. Croft, 1995; Steedman, 1991; Vallduvi, 1992).

Summary

The conclusion warranted by the studies we have reviewed here is that the presence of prosodic information cueing a boundary, or of a sentence accent, can have an effect on syntactic analysis; it can lead the listener to prefer analyses that are consistent with the prosodic information provided. However, these effects are far from robust and determinate, and there is little evidence for early exploitation of prosodic information in processing. The strongest effects that have been observed are those associated with boundaries; but note that this can be viewed from two angles — as prosodic signaling of syntactic breaks, or as prosodic signaling of grouping (i.e., of cohesion where there is no break). These two views are of course not independent, but they do correspond to somewhat different questions from the point of view of the researcher studying syntactic processing. The effect of noting a break will be to close off a syntactic constituent, while the effect of noting cohesion will be to add a node to the current constituent. The studies reviewed above have not always been explicit about the exact role of prosodic information in the parsing process.

Many of the studies we have described have produced results which are consistent with listener failure to exploit available prosodic information. It is clear that frustration has met the expectations of any researchers who proposed that syntactic and prosodic hierarchies

would be isomorphic, with prosodic information cueing syntactic information directly. The prosodic hierarchy as defined by, for example, Selkirk (1984) or Nespor and Vogel (1986) is flatter than the syntactic hierarchy, and the labeling of one hierarchy does not map neatly onto the labeling of the other (Shatruck-Hufnagel & Turk, 1996).

In general, we feel that although there has been a veritable explosion of interest in the relationship of prosody and syntactic processing in the last few years, the work so far has not led to clear theoretical advance. There has been in our view too little discussion of the relative time-course of prosodic processing, lexical processing, and syntactic processing in sentence comprehension, and we know of no attempts to implement computational models of this interaction. There has also in general been regrettably little attention paid to characterizing the acoustic dimensions making up the prosodic information putatively involved in many observed effects, as well as to establishing the range of possible parsing effects of prosody (although see Nicol, 1996, for consideration of this issue). Prosodic cues to the presence of a boundary have been the most reliable source of significant effects on parsing decisions (although see Tyler & Warren, 1987, for an argument that prosody signals syntactic cohesion; see Shillcock, Bard, & Spensley (1988) for an argument that prosodic grouping facilitates word recognition; and see Scott and Cutler, 1984, for evidence that cohesion can be signaled by phonological assimilation which would certainly be inhibited by a prosodic boundary). However, boundary signals have also been the most studied phenomena. Thus it is too early to conclude that prosody can cue breaks (construction of new constituents) but not cohesion (elaboration of current constituents), although the evidence so far might be consistent with such a claim.

If such a conclusion does eventually prove warranted, one underlying reason may be the range of alternative prosodic realizations which any given structure may receive; only relatively unambiguous signals may be exploited. Thus it may be that the parser operates on the assumption that a prosodic break is highly unlikely to map onto syntactic cohesion, while prosodic cohesion may well map onto a syntactic break. It is unfortunate that so few studies (Grabe et al., 1995, being a welcome exception here) have compared alternative possible prosodic realizations of a given syntactic structure (see Ferreira et al., 1996, for similar remarks). Prosodic structure, as we have repeatedly pointed out, is rarely fully determinate; it allows for variation which is both qualitative (alternative patterns which are acceptable) and quantitative (more vs. less strongly realized markers). Murray et al. (submitted) speculate that stronger processing effects of prosody with a given syntactic ambiguity in some studies than in others may be due to differential choice of stronger versus less strong realizations of the prosodic disambiguation; but in the absence of agreed standards for comparing prosodic realizations this suggestion can hardly be tested.

One obvious point is that the manner in which prosodic information is exploited is language-specific. To take an extreme example, English listeners have been shown to (mis)interpret the long closure of Japanese geminate stops as pauses, and consequently to postulate strong syntactic boundaries occurring at what are actually word-internal consonant sequences (Beckman, 1996). Although Bolinger (1978) has termed boundary signaling a true prosodic universal, and despite cross-language similarities in prosodic correlates of syntactic structure (Vaissiere, 1983), languages nevertheless differ in the permissible mappings of prosodic to syntactic structures. The effects of such language-specificity have so far not been charted.

What does seem to us a welcome development is that recent perceptual studies have attempted to be more precise in the definition of what they assume relevant prosodic information to be, and to take into account the insights arising from recent developments in phonological theory. Greater phonetic precision, more consideration of the implications of prosodic ambiguity for the exploitability of prosodic information in parsing, consideration of cross-language variation, and above all a theoretical framework allowing explicit prediction of processing effects, will all be of great value to future work.

PROSODY IN THE COMPREHENSION OF DISCOURSE STRUCTURE

Introduction

Just as prosody is closely related to syntactic structure, so is it closely connected with discourse structure. Bolinger (1978) lists besides the universal relationship of prosody to syntactic breaks a second prosodic universal: the highlighting of salient information. In Bolinger (1972) he used the term "point of information focus" for sentence accent. In this focus-based account of accentual structure, whether or not a word is prominent depends both on its relative importance within the sentence and also on the context of the sentence itself. To take an obvious English example, the sentence *George has flowers for Mary* would be an appropriate response to any of the following questions: 1) *Who has flowers for Mary?*, 2) *What does George have for Mary?*, 3) *Who does George have flowers for?* (example from Eady and Cooper, 1986). However, a different noun would be focused in each case: *George* for the first question, *flowers* for the second, *Mary* for the third. If the sentence were spoken, this focus should be realized acoustically, usually via an increase in FO or in FO movement, intensity and/or duration of the focused word. Depending on the semantic/pragmatic context, different prosodic structures are appropriate.

Just as the relationship between prosodic and syntactic information has both some language-universal and some language-specific characteristics, so does the relationship of prosody to discourse in part differ across languages. The same differences in focus which result in differences in accent placement in English may for instance be expressed in differences in word order in languages such as Italian or Catalan (see Ladd, 1996; Vallduvi, 1992). Similarly in French (as in English), focus can be expressed by differences in word order, in particular via clefted syntactic construction and topicalization by dislocation (e.g., *C 'est le set, que je t'ai demande* "It's the salt I asked you for;" Perrot, 1978; Dahan, 1994; Dahan & Bernard, 1996). But apparent differences across languages may not always be clear-cut. In any language, word-order manipulation for focus will result in rearrangement of the prosodic structure as well. Still, there are also more clear-cut language-specific effects in how focus relates to prosody. There is evidence, for instance, that in languages such as Korean and Japanese, in which the intonationally determined pitch accents such as are found in English and Italian do not occur, local pitch range expansion and dephrasing play functionally analogous roles to accenting and deaccenting (Venditti, Jun, & Beckman, 1996). The fact that the remainder of this section deals nearly exclusively with studies which have been carried out in English and in Dutch, and that in these two languages the relationship between prosody and discourse appears to be realized in the same manner, should not be taken as an indicator that the observed processing effects may be encountered in all languages. The

review merely reflects the type of work that has predominated and has so far reached the psycholinguistic journals; similar studies in other languages are plainly needed.

Accenting and its relationship to the computation of information structure have been the subject of a large body of work, which accounts for about half of the material to be reviewed in this section (the following two subsections). Studies of the role of prosody in the determination of reference, and of topic structure in discourse, plus a small set of findings on conversational structure, are combined under the general heading of construction of a discourse model (the last subsection).

Accent, focus, and the prediction of accent

Words which bear sentence accent are processed more efficiently than words which do not, as measured for instance by the speed with which a word-initial phoneme can be detected (Shields, McHugh, & Martin, 1974; Cutler & Foss, 1977) or a mispronunciation in the word can be registered (Cole, Jakimik, & Cooper, 1978; Cole & Jakimik, 1980). This finding can in principle be explained by heightened acoustic clarity of stressed syllables, which facilitates the lexical and/or segmental processing of these speech events. However, this is clearly not all that is going on. The phoneme-monitoring response time advantage of accented words is not solely due to acoustic factors. Cutler (1976) recorded sentences in two prosodic versions, one in which the target-bearing word bore contrastive accent (e.g., *That summer four years ago I ate roast DUCK for the first time)* and one in which contrastive accent fell elsewhere (e.g., *That summer four years ago I ate roast duck for EVERY MEAL).* The target-bearing word itself (i.e., *duck)* was then edited out of each version and replaced by acoustically identical copies of the same word taken from a third recording of the same sentence, in which no contrastive accents occurred. This resulted in two versions of each experimental sentence, with acoustically identical target-bearing words but different prosodic contours on the words preceding the target: in one case the prosody was consistent with sentence accent occurring at the location of the target, in the other case it was consistent with accent falling elsewhere. Under these circumstances the target in the "accented" position was still responded to significantly faster than the target in the "unaccented" position. Since there were no acoustic differences between the target words themselves that could account for this result, and the only difference in the preceding context lay in the prosody, Cutler concluded that listeners had been using prosodic information to predict where accent would occur. For Dutch, Nooteboom (1995) suggested that "flat hat" intonation patterns (see 't Hart, Collier, & Cohen, 1990) may be especially useful in allowing listeners to look ahead and predict accents.

Subsequent studies examined the components of the prosodic pattern which contributed to this effect. When pitch variation was removed—that is, the sentences were monotonized—the predicted accent effect was unchanged; and it was also not affected by manipulation of the duration of closure for the target stop consonant (Cutler & Darwin, 1981). Thus the effect appears to be robust and, Cutler and Darwin argued, probably not crucially dependent on any particular prosodic dimension. However, when speech hybridization techniques were used to interchange timing patterns between the two versions of an utterance such as the example above *(That summer four years ago I ate roast DUCK for the first time* vs. *That summer four years ago I ate roast duck for EVERY MEAL),* so that "impossible" utterances resulted (e.g., an utterance in which the rhythm suggested that accent would fall on the target-

bearing word while the FO contour suggested that it would fall elsewhere), the predicted accent effect disappeared (Cutler, 1987) suggesting that consistency among the separate suprasegmental dimensions is important for listeners to be able to exploit them efficiently.

A rather more general proposal concerning the predictive use of prosodic information, however, was put forward by Shields et al. (1974) on the basis of Martin's (1972) rhythmic-structure theory, according to which the rhythmic events of speech are hierarchically organized. This organization, Shields et al. argued, makes the timing of speech events predictable; temporal redundancy in the speech can be exploited in processing. In Shields et al.'s phoneme-monitoring study, listeners heard nonsense words embedded in real sentences. Detection of the initial phoneme of the nonsense word was faster when the first (target-bearing) syllable was stressed rather than unstressed; but importantly, the effect disappeared when the nonsense word was embedded in a string of other nonsense words, suggesting that the facilitation was not due simply to acoustic advantage. The authors concluded that listeners can predict upcoming stresses from the preceding rhythmic structure.

Other studies supported this predictive view, by showing that the disruption of rhythm impairs performance on many perceptual tasks. Martin (1979), for example, found that either lengthening or shortening a single vowel in a recorded utterance could cause a perceptible momentary alteration in tempo, and increase listeners' phoneme-monitoring response times. Meltzer et al. (1976) found that phoneme targets which were slightly displaced from their position in normal speech (by deleting 100ms of the material immediately preceding the target phoneme) were detected more slowly (e.g., *He laughed and laughed till hi-* /-100ms deleted-/ *belly wiggled like jelly).* Buxton (1983) found that adding or removing a syllable on a word preceding a phoneme target (replacing *blue* by *reddish,* or *blueish* by *red)* also increased detection time (in comparison to manipulations which resulted in no change to the number of syllables, such as replacing *blueish* by *reddish,* or *blue* by *red).*

All these results suggest that listeners process a regular rhythm, using it to make predictions about temporal patterns; when manipulations of the speech signal cause these predictions to be proven wrong, recognition is momentarily disrupted. However, more recent results have called into question the interpretation proposed for those findings. Mens and Povel (1986) conducted an experiment modeled on those of Meltzer et al. (1976) and Buxton (1983), but avoiding any segmental disruption. The most important temporal modifications were brought about by replacing the pretarget word by one of a different number of syllables (e.g. *kat* "cat" was replaced by *kandidaat* "candidate"). Mens and Povel failed to replicate the predictability effect of rhythm. Pitt and Samuel (1990) similarly only weakly replicated Shields et al.'s (1974) stress-predictability effect, using acoustically controlled target-bearing words embedded in a natural sentence context; strong predictability effects only occurred when the word was embedded in a rhythmically highly regular list of words. These authors speculated that natural sentence contexts may offer little opportunity for exercising prediction with respect to the location of stressed syllables. Tyler and Warren (1987), however, showed that disruption of local prosodic structure did lead to longer monitoring latencies, and speculated that such disruption will adversely affect processing only if it interferes with computation of the prosodic structure (the hierarchy of phonological phrasing) of the utterance. This suggests that predictability effects might be constrained by the prosodic structure, but this might only rarely be such as to produce the sustained regularity which,

according to Pitt and Samuel, listeners need if they are to exploit the predictability.

It will be noticed that different facets of prosodic structure have been at issue in this research. Shields et al.'s hypothesis concerns the level of utterance rhythm: in this case, the pattern of stressed and unstressed syllables. But in their experiment, this variable was actually confounded with the factor of sentence accent. Both in their study and in that of Meltzer et al. (1976), the nonsense words seem to have been carrying the main information of the clause in which they occurred, so that the speaker would presumably have assigned them sentence accent. The studies of Cutler (1976), Cutler and Foss (1977) and Cutler and Darwin (1981) all manipulated sentence accent. The strong and consistent effects of sentence accent patterns, combined with the fragility of predictability effects per se, raise the further possibility that predictability effects attributed to rhythmic regularity could actually be due to the active search by listeners for focus information. Further experiments are probably desirable, to disentangle the effects of syllable-level stress on the one hand, and accent, applied to words in sentences, on the other, and to examine these effects explicitly with respect to the phonological-phrasing structure.

Certainly there does appear to be active search for sentence focus; Cutler and Fodor (1979) showed that semantic focussing leads to faster responses in phoneme monitoring in just the same way as prosodic accentuation does. Thus when listeners have determined where accent falls, they have located the focussed or informationally prominent part of an utterance; an active search for accent may therefore represent an active search for the semantically most central portion of a speaker's message. Focussed parts of words receive more detailed semantic processing: Multiple meanings of homophones are activated if the words are in focus, but not necessarily if the words are not in focus (Blutner & Sommer, 1988). Retention of the surface form of a word in memory is more likely if the word was in focus in a heard sentence than if it was not (Birch & Garnsey, 1995). In a study by Louise Lee Seng (referred to by Cutler, 1982) both semantic focus and prosodic focus were investigated. The effects of semantic focus and prosodic focus turned out to be significant, equally strong, and additive. This suggests that the search for sentence accent on the basis of prosodic information and the search for semantic focus on the basis of nonprosodic information could be seen as separate strategies that proceed in parallel but have the same goal.

Accenting, de-accenting, and information structure

One reason why speakers mark items as focused or not is frequently said to be the "newness" versus "givenness" of the items (Chafe, 1974; Brown, 1983; Nooteboom & Terken, 1982) — although definitions of what is given and what is new vary considerably (Halliday, 1967; Chafe, 1974; Clark & Haviland, 1977; Prince, 1981). The results of Read et al.'s (1980) study, described above, in which globally ambiguous who-questions such as *Wie zoent de vrouw?* were disambiguated by accent placement, were interpreted by the authors in terms of information structure; accent on the verb effectively deaccented the following noun, implying that the noun is an existing topic of the discourse, which implies further that it is the grammatical subject of the sentence, and the question word thus the grammatical object. Contrastive accent on a word leads to rapid and efficient recognition of that word, as the research reviewed in the preceding subsection showed, and the presence of contrastive accent is also rapidly exploited to derive information about the sentence semantics. Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard, and Carlson (1995) asked listeners to select one

of several items in a display, such as a large red square from a set of four items comprising a large red square, a large blue circle, a small red square, and a small yellow triangle; in this instance, correct selection was possible on hearing the word *large* in *the LARGE red square,* suggesting that the contrastive accent allowed the listener immediately to select the one member of the set of large items which contrasted on exactly that dimension with some other item.

The relation between accent patterns and discourse structure suggests that overall sentence-processing measures should in general be sensitive to accent placement, and a substantial number of sentence-comprehension studies have manipulated this factor directly. Listeners' judgments of prosodic appropriateness are higher when new information is accented and old information is not (Birch & Clifton, 1995). Thus it is not surprising to find that response time in a simple comprehension task is also shorter when new information is accented and given information is not, compared to conditions in which the accent pattern is reversed (Bock & Mazzella, 1983). Although accenting of given information is judged more acceptable by listeners than deaccenting of new information (Nooteboom & Kruyt, 1987), the presence of an accent on given information seems to affect sentence comprehension. Terken and Nooteboom (1987) found that subjects' response times to judge whether a spoken sentence (e.g., *the P is on the left of the K)* correctly described a visual display were facilitated by appropriate accent placement—newly mentioned entities should be accented and entities that had already been mentioned should be deaccented. Inappropriate accentuation on repeated words slowed response times. Thus it was not simply the case that more salient words tended to receive more processing attention; the accent structure has to have been processed with respect to the sentence semantics. Consistent with this view of accent processing, Birch and Clifton (1995) found that acceptability judgments for simple question-answer pairs *(Why is Ken smiling? He won the lottery)* were equally fast when focus on a verb phrase was signaled narrowly via accent on a noun *(He won the LOTTERY)* or broadly via accent on verb and noun *(He WON the LOTTERY);* although there were two accents in one version and only one in the other, the focus structure was identical in both.

Donselaar and Lentz (1994) observed that the degree of speech intelligibility influences listeners' use of this interdependence between information and accent structure. A forced-choice task with target words embedded in sentence contexts was employed in their experiments. Both normal-hearing and hearing-impaired subjects were tested; they heard pairs of questions and answers. The target words in the answers referred either to given or new information and were either accented or unaccented. For example, after a question containing the word pop ("doll"), the answer contained an accented or unaccented realization of either *pop* (now given information) or *pot* ("pot;" new information). After each question and answer, two words appeared on a visual display (for this example the words would be POP and POT) and the subjects had to decide which of these two words they had just heard in the answer. When speech quality was relatively high, normal-hearing subjects answered more accurately for accented than unaccented words, regardless of their information value. Hearing-impaired subjects relied more strongly on the correspondence between prosody and information value; their response accuracy dropped sharply when given words were accented, suggesting that they interpreted accented target words as being new. However, when the segmental quality of speech was impaired, normal-hearing listeners showed this pattern, indicative of an interaction between information and accentuation, as well.

In these experiments, information status was defined in terms of the number of word occurrences: The first occurrence of a word was assumed to convey new information, the second occurrence given information. A study by Terken and Hirschberg (1994) suggests, however, that simple prior mention may not suffice to motivate deaccentuation. In their study subjects described a visual display of objects such as a ball, a cone, a cross, a diamond, and so forth. After a context in which a ball had been the topic of conversation *(the ball touches the cone, the ball touches the cross, the ball touches the diamond)* speakers were more likely to deaccent ball in the description of a *ball touches star* event than in that of a *star touches ball* event; that is, deaccentuation occurred more frequently if the grammatical role and surface position of the given information were *both* repeated. Prior mention in any form was not enough. Terken and Hirschberg proposed that the properties of grammatical function and surface position may be used by listeners as cues to access potential antecedents in the discourse model. They assumed that reduced accessibility may be signaled by the accenting of the target expression. Deaccentuation of targets with identical grammatical roles and surface positions, and accentuation of other targets, thus contribute to the pruning of the set of candidate antecedents. Needham (1990) similarly showed that deaccentuation can occur on a word which refers to just a part of an object already mentioned in the previous discourse context, but only if the part is central to the object; words referring to peripherical parts will tend to be accented.

Subsequent reference to a previously-mentioned concept via a different surface form has been relatively little explored in studies of prosodic processing. However, a study by Donselaar (1995a) compared the effects of accentuation of synonymous references (e.g., *ship-boat)* versus identical references (e.g., *boat-boat)* in Dutch. In an auditory verification experiment (in which subjects make true /false judgments about spoken sentences), she first presented subjects with utterance pairs such as: *The multimillionaire bought a surprise for his wife. He gave her a boat/ship/mink.* Immediately afterwards they had to verify a third utterance, for example, *The wife UNEXPECTEDLY got a BOAT/boat.* The target word (e.g., BOAT) in the third utterance was either accented or not, and accents were realized in a localized manner, as so-called pointed hats (see 't Hart, Collier & Cohen, 1990) in order to minimize acoustic differences across versions. Sentences with synonymous target words *(boat* after a *ship* context) were verified less rapidly than sentences with identical target words *(boat* after a *boat* context). However, unaccented synonyms were verified significantly faster than accented synonyms. There was no difference in the verification latencies for accented versus unaccented identical words. This finding suggests that prosodic information is more extensively used in resolving synonymous references than identical references. However, the use here of synonyms for subsequent reference via a different surface form (rather than, for instance, superordinate concepts) may have rendered the materials unnatural, and it is not clear whether listeners would have interpreted the utterance to be verified as continuous with the preceding discourse or not. The use of prosody in resolving different types of anaphors clearly needs further investigation.

So far, it has been implied in the research described that listeners use a binary classification of words as either being accented or unaccented. Fowler and Housum (1987) did not use such a binary accent classification, but instead manipulated the operational notion of intelligibility. They observed that first occurrences of words in a radio monologue were acoustically longer and, in a simple out-of-context recognition task, proved to be

more intelligible than second occurrences. Bard, Lowe, and Altmann (1989) also showed that second tokens of a word are less intelligible than their corresponding first tokens. The presence of an accent on the first occurrence of a word could of course account for its higher intelligibility. Not only does the presence of accent increase word duration (Klatt, 1976; van Santen & Olive, 1990; Eefting, 1991; Dahan & Bernard, 1996), it also leads to greater spectral clarity (Koopmans-van Beinum & van Bergem, 1989). In fact, it was shown by Hawkins and Warren (1994) in identification experiments on words and CV segments excised from conversational speech that sentence accent exerts a greater influence on intelligibility than whether or not the word has been used before in the conversation.

Variation in intelligibility for first versus second mentions of words could result either from an on-line assessment of redundancy by the talker, or from the mere repetition of the same word. Fowler (1988) addressed this question and found that repeated productions of homophones in their different readings (e.g., *right* followed later by *write)* did not yield shortening, so that the effect could not simply be ascribed to articulatory routine. Repetitions of words in a list did not lead to shortening either. Likewise, Bard et al. (1989) failed to find shortening of repeated words if the words were not coreferential. It appeared that shortening only occurred for repetitions of a referent, only in meaningful prose, and to a larger extent if this was produced spontaneously in a communicative setting. The shortening thus may reflect the talker's estimate that a listener has sufficient other information to identify the word. This corresponds with Lieberman's (1963) and Hunnicutt's (1985) findings that the intelligibility of a word is higher when this word has been produced in a less informative context. Bard, Sotillo, Anderson, Doherty-Sneddon, and Newlands (1995) however questioned this implication of cooperative behavior on the part of the speaker; they showed that a speaker reduces the intelligibility of a word when it has been referred to in the preceding context, regardless of the listener's knowledge about an earlier mention.

Given the close relationships between accentuation, intelligibility, and the given or new information that words convey, does the intelligibility of words have an effect on the way the semantic content of the utterance is processed? Some evidence suggests that the degree of intelligibility of words influences the way those words are related to previous discourse. Fowler and Housum (1987) showed that listeners could identify first and second productions as such. The supposedly less intelligible second productions proved to be better recognition cues (as measured by the speed with which a recognition response was provided) in a test of listeners' memory for whether a word had occurred in the narrative. This was not just the case for the word itself (an effect which in this study could have reflected the durational differences between reduced second productions and unreduced first productions), but also for words close to it in the text. The less intelligible words apparently aided listeners to refer back to earlier presented information. Bard, Cooper, Kowtko, and Brew (1991) replicated and extended the probe-recognition experiment by Fowler and Housum, but failed to achieve the same results. They attributed this to a difference in speech materials: the dictations they employed had more interruptions, disfluencies, self-corrections, and so forth, than the broadcast materials used by Fowler and Housum, and the differences in intelligibility between first and second occurrences were perhaps too small to rise above all the naturally occurring variation in the materials. In a subsequent experiment, Bard et al. reduced the intelligibility of the prime words by presenting them at lower signal-to-noise ratios. The underlying idea was that words that were less intelligible would prove to

be better primers of earlier materials. This appeared to be true, but there was only a significant priming advantage for less intelligible over more intelligible primes when prime and target were coreferential.

Fowler and Housum (1987) and Bard et al. (1991) suggest that the intelligibility of words determines whether words are processed as referring to given or new information. Talkers generally reduce the intelligibility of repeated words, and this helps listeners in that it provides an indication that these words refer back to earlier-presented information in the discourse. How this (reduction in) acoustic information is processed, and the discourse information extracted by the listener, remain, however, unclear. It is possible that relatively intelligible information is easy to process, and there is little need to rely on contextual cues to recognize the incoming information. Relatively unintelligible information is more difficult to process, and preceding contextual information is needed. Such words would "automatically" be related to previous concepts, simply because invocation of these concepts would be needed to help the words be recognized. The connection with earlier given information in the discourse model, in other words, is not triggered by any specific mechanism, but is simply a by-product of the search for additional information to process less intelligible words. Bard et al. (1991) propose that the unintelligibility of words is thus, in effect, translated into their power to reactivate related material.

Terken and Nooteboom (1987) suggested a more specific hypothesis, namely that the presence of an accent leads listeners to attend to acoustic/phonetic properties of the signal, since a new interpretation must be constructed, while lack of accent leads them to seek an interpretation in the set of already activated discourse entities, with less consequent need to attend to details of the input. Donselaar (1991,1995b) tested this suggestion by measuring listeners' attention for the acoustic/phonetic details of accented versus unaccented "given" and "new" words (defined as second vs. first occurrences) by means of mispronunciation detection; her experiments failed to find evidence for Terken and Nooteboom's hypothesis. Her results showed that the presence of accent only facilitated the detection of mispronunciations in new (i.e., first occurrences of) words, whereas mispronunciations in accented and unaccented given words (second occurrences) were detected equally fast. In a subsequent study (Donselaar, 1995b) subjects detected clicks in accented versus unaccented "given" and "new" words; again, response times showed an effect of information structure only: detections were faster in new than in given words. When the same materials were used in an off-line click localization experiment, however, an interaction between information and accent structure appeared: more correct localizations were made for clicks in accented new and unaccented given words, than in unaccented new or accented given words. The two click tasks varied only with respect to when subjects' responses were tapped: early in click detection, late in click localization. Donselaar concluded that there is no evidence for an early role of prosody in the sense that Terken and Nooteboom suggested, namely that the presence versus absence of accents can trigger predominantly bottom-up versus top-down processing; she doubted further whether Terken and Nooteboom's sentence verification paradigm taps these early stages of processing.

The construction of a discourse model

The research reviewed in the preceding section suggests that the acoustic/prosodic realization of a word greatly influences the way the concept it conveys is integrated into discourse

structure. The acoustic realization will determine whether the word should be linked to a previously-mentioned concept, or will be assumed to be introducing a new concept. The reduced intelligibility of words which have previously been referred to, as manipulated in the studies just described, may in fact serve a role similar to the role of pronoun anaphors; it has been assumed that pronouns access the conceptual representation of their antecedent directly, whereas nominal anaphors (repetition of a previously mentioned noun) initially prime a surface form of representation as a preliminary to accessing associated conceptual information (Cloitre & Bever, 1988). The parallel between given information and anaphoric expression is supported by similarity between Fowler and Housum's (1987) results and the demonstration by McKoon and Ratcliff (1980) that anaphoric expressions can effectively prime syntactically related items. Unaccented words, or more generally words which have reduced intelligibility, could thus be processed in the same way as anaphoric devices in discourse structure integration. Indeed, Hirschberg and Ward (1991) studied the effect of accenting versus deaccenting anaphoric expressions (e.g., in *Mary said [SHE/she] deserves the scholarship and so did Cathy)* on listeners' interpretation of the binding of the anaphor; they found that deaccented anaphors were more likely to be assigned the reading *Cathy said Mary deserves the scholarship.* Analogously, a nonce lexical anaphor such as *butcher* in *The doctor bungled the operation and then had the audacity to charge $10,000. We should sue the butcher* (example adapted from Ladd, 1980) must be deaccented to make the anaphora clear; accentuation leads to a completely different (and ridiculous) pragmatic interpretation. In this way the prosodic/acoustic characteristics of words can be of direct use in the semantic/pragmatic analysis of an utterance, and the processing of prosodic information can be seen to be integrated into the processing of linguistic structure at the discourse level as we have seen it to be at the lexical and syntactic levels. (Note that this claim is similar to that put forward by Schafer et al., 1996, described above in the discussion of the processing of local syntactic ambiguity.)

Gernsbacher and Jescheniak (1995) proposed that forward reference also draws directly upon prosodic processing. Using a probe recognition task with short narratives, similar to the designs used by McKoon and Ratcliff (1980) and Fowler and Housum (1987), Gernsbacher and Jescheniak investigated accentuation as a cataphoric (forward reference) device. Accentuation facilitated responses to probes presented immediately after the accented word, and also to the same probes presented later in the narrative; moreover, responses to the same probes were inhibited when some other prior word had been accented. Another cataphoric device, the indefinite *this,* had a similar (though somewhat weaker) effect to that of accent. Gernsbacher and Jescheniak proposed that key concepts—those concepts that play a central role in discourse — are marked with cataphoric devices. By virtue of this marking, these key concepts gain a privileged status in the mental representation that listeners build when comprehending discourse; for instance, they are protected from being suppressed by subsequently mentioned concepts. Gernsbacher and Jescheniak suggested that cataphoric devices, like anaphors, trigger mechanisms of enhancement and suppression to improve their concepts' accessibility in the discourse model constructed by a listener in the processing of a narrative.

It has also been proposed that prosody can help to signal the presence of an implicit anaphor, for example, a gap marked by the presence in the sentence of a Wh-word; note that explicit and implicit anaphors relate to their antecedents in the same way (Bever & McElree,

1988). A study by Nagel, Shapiro, and Nawy (1994) used wh-questions in which the gap location was varied: *Which doctor did the supervisor call__to get help for his youngest daughter?* versus *Which doctor did the supervisor call to get help for during the crisis?* In the first question, there is a gap that must be coindexed with *doctor* after *call*, but in the second question there is no gap after *call.* An acoustic analysis identified prosodic correlates of this variation: The verb *call* was longer in the first question than in the second question, and in the first question there was also an FO fall on the same word. These sentences were used in a cross-modal priming study, in which listeners performed a lexical decision on a related probe (e.g., PATIENT), unrelated probe (e.g., CURRENT) or pseudoword (e.g., COMTENT) presented after *call.* In the the first question, with the gap after *call,* reaction times to probes related to the antecedent (PATIENT-doctor) were significantly faster than RTs to unrelated controls. In the second question, with no gap, no such facilitation was found. This suggests, according to Nagel et al., that prosodic information is used on-line to decide whether or not to posit a gap at possible gap locations, such as following *call.* However, the response-time effect was actually due to long latencies for the control words in the gap condition (rather than faster latencies for the related words); Nagel et al. account for this by proposing that prosodic signals of gap location trigger recruitment of increased processing resources.

Considerable research has also been carried out on the prosodic correlates of the division of discourse into topics. Most of this research has involved production studies. Speakers start new topics relatively high in their pitch range and finish topics by compressing their range (Brown, Currie, & Kenworthy, 1980; see also Venditti & Swerts, 1996); they use low boundary tones at topic endings but not for continuations (Swerts & Geluykens, 1994). Similar reports of the role of FO declination in structuring discourse may be found in Bruce (1982), Menn and Boyce (1982), Thorsen (1985), Ladd (1988), and Sluijter and Terken (1993). Speakers pause at "paragraph" boundaries in re-telling a story (Van Donzel & Koopmans-van Beinum, 1996); these boundaries are more likely to be marked by a filled pause (Swerts, Wichmann, & Beun, 1996); and the duration of pauses between utterances is longer for major than for minor topic shifts (Lehiste, 1979; Swerts & Geluykens, 1994; Brown et al., 1980). Speaking rate is also associated with text structure: the rate is lower utterance- or paragraph-initially than -finally (Brubaker, 1972; Lehiste, 1980; Butterworth, 1975). Finally, Brown et al. (1980) found that there was a rise in amplitude at the beginning of a topic, and a fall at the end.

Thus speakers apparently provide ample prosodic information on the topic structure of discourse. However, the use by listeners of prosodic cues in structuring discourse has not been as extensively investigated as the production phenomena. A series of studies by Hirschberg and colleagues studied the relationship between particular intonational phenomena (phrasing, accent placement, pitch range) and discourse using both acoustic analyses of spoken utterances and listener evaluation of the discourse attributes of the same utterances. Hirschberg & Pierrehumbert (1986) proposed that a hierarchical segmentation of discourse can be marked by systematic variation in pitch range, which signals movement between levels in the segment hierarchy. Major boundaries are marked by the largest increases in pitch range, whereas smaller increases denote subsegment boundaries. The amount of final raising or lowering at the end of the phrase indicates the degree of conceptual continuity between phrases. Variation of final lowering can also mark the internal structure of discourse

segments. Hirschberg and Pierrehumbert proposed, furthermore, that the assignment of pitch range at the discourse-segment boundary can enforce one segmentation of a given discourse over another and can disambiguate among potentially ambiguous reference resolutions. While an increase in the pitch range indicates discourse-segment boundaries, a reduction in the amount of final lowering at a potential boundary can indicate that no such boundary is in fact present—that is, can indicate that a given utterance and one which follows it are part of the same discourse segment.

Pierrehumbert and Hirschberg (1990) observed that speakers choose among a repertoire of prosodic means—pitch accent, phrase accent, boundary tone—to convey specific types of relationships between the current utterance and previous and subsequent utterances. Pitch accents are used to convey information about the status of discourse referents, modifiers, predicates, and relationships specified by accented lexical items. Phrase accents, on the other hand, are used to convey information about the relatedness of intermediate phrases. Finally, boundary tones convey information about whether the current phrase is "forward looking" or not.

Grosz and Hirschberg (1992) and Grosz, Hirschberg, and Nakatani (1994) conducted corpus-based empirical work on intonational features of spoken language. They studied the relationship between acoustic-prosodic variation and discourse structure, as determined from an independent model of discourse. This model, by Grosz and Sidner (1986), distinguishes between two levels of discourse processing: global and local. The global level concerns discourse segments (topics), plus their embedding structure and other relations. The local level concerns features of the utterances within a discourse segment and relations among these. Grosz and collaborators examined the relations between prosodic features and discourse structure of three wire service news stories. Discourse structure was labeled by subjects either from text alone, or from text plus speech. They found statistically significant associations of aspects of pitch range, amplitude, and timing with features of global and local structure available from the text alone (parenthetical remarks use expanded pitch range, for example). Moreover, they found that the judgments of listeners who heard the spoken utterances correlated reliably with these acoustic and prosodic features, suggesting that the listeners had used the prosodic information in making their discourse decisions.

Hirschberg and Nakatani (1996) explored a corpus of direction-giving monologues and examined the effects of speaking style (spontaneous vs. read) and of discourse segmentation method (based on the text alone, or on the text with the spoken version). Although they reported results from a single speaker only, their data seem to indicate that their method for discourse analysis — based on Grosz and Sidner (1986)—provides reliable segmentations of spontaneous as well as read speech. The availability of speech led to higher reliability scores, indicating that the labelers made use of prosody. Hirschberg and Nakatani also compared the acoustic-prosodic features of initial, medial, and final utterances in a discourse segment. The segment-initial utterances differed from medial and final utterances in both prominence and rhythmic properties. Segment-medial and segment-final utterances were distinguished more clearly by rhythmic features, primarily pause. The listeners in the labeling task were presumably exploiting these prosodic realizations to increase their accuracy.

Some further evidence is available from simple paragraph-structure studies. Sluijter and Terken (1993) had two speakers read three-paragraph texts in versions which differed

as to where in the text a given target sentence occurred: at initial, medial, or final position in the second paragraph. Differences in the intonation of the target sentences could only be attributed to the position of these sentences in the text. The target sentences were then excised from the recordings and presented to listeners who had to decide which position these sentences had originally occupied in the paragraph. The listeners could, in general, assign the sentences to the correct original positions, although judgments were rather less accurate for sentences from medial positions. Acoustic analyses suggested that the speakers performed some form of supradeclination: there was sequential lowering of baseline and topline onsets in the course of the paragraph, on which listeners could potentially base their decision. In the studies by Lehiste and Wang (1977) and Kreiman (1982) referred to in the preceding section, listeners judged paragraph as well as sentence boundaries, with a fair measure of success, though performance was better in the sentence than in the paragraph condition.

The perceptual use of the prosodic correlates of topic structure in discourse also formed part of the study of Swerts and Geluykens (1994) of which the production results were described above. When utterances were band-pass filtered to remove information about content, listeners successfully employed melodic and (to a somewhat lesser extent) pausal information to process the signal in terms of discourse structure. More recently, Swerts (1997) investigated a corpus of twelve spontaneous Dutch monologues. He introduced a method of experimentally determining hierarchical discourse boundaries by computing the proportion of subjects agreeing on a given break. Acoustic measurements of the speech materials showed that prosodic variables such as pause length, pitch reset, and the proportion of low boundary tones increase continuously with perceived boundary strength at the discourse level. Evidence for an extensive use of melodic information was also found by Silverman (1987). In the study of paragraph structure by Kreiman (1982), listeners' judgments appeared not to be based on pause information, but rather on acoustic characteristics of the onset of the following utterance (increase in FO and amplitude).

Listeners may not rely on pause structure because it is an unreliable boundary cue in spontaneous speech. In comparison with read or rehearsed speech, spontaneous speech contains longer and more frequent pauses and hesitations (Barik, 1977; Crystal & Davy, 1969; Kowal, O'Connell, O'Brien, & Bryant, 1975; Levin, Schaffer, & Snow, 1982), and shorter prosodic units (Crystal & Davy, 1969). Blaauw (1994) reports that 21% of all pauses in a large sample of spontaneous speech were associated neither with syntactic structure nor with a prosodic structure derived from syllabic weights and accents (see also Henderson, Goldman-Eisler, & Skarbek, 1966, and Levin et al., 1982, for further reports of the lack of parity between pause structure and syntactic structure in this type of speech). Blaauw suggests that pauses in spontaneous speech may be related to information structure; they tend to occur prior to highly informative words (again, this is a suggestion with considerable history in the pausing literature: Maclay & Osgood, 1959; Goldman-Eisler, 1968). As Hirschberg (1993) pointed out, most work on discourse has defined discourse as "utterances in context"; monologues, elicited speech, read speech, and radio speech have been more frequently examined than natural dialogue. It is questionable to what extent the results of such studies can be generalized to processing in natural discourse situations. Mehta and Cutler (1988) found that phoneme-detection effects that are related to prosody, such as faster detection of targets in accented than in unaccented words, and in strong than in weak syllables, emerged with spontaneously spoken, but not with read materials. Thus

studies of the on-line comprehension of natural discourse are needed before firm conclusions can be drawn about listeners' processing of its structure.

At a higher level in discourse structure, the use of prosodic cues in turn-taking was examined by Cutler and Pearson (1986). They analyzed prosodic cues in short dialogues that were read aloud by two speakers. The same utterances occurred at turn-final or turn-medial position in different versions of the texts. These utterances were presented to listeners, in isolation and pairwise. Results indicated that listeners did not distinguish between utterances being turn-final or not, but some utterances were mainly judged turn-final and others turn-medial. Prosodic transcriptions of these utterances revealed that turn-final judgments were associated with downstepped contours on the final tone groups, while turn-medial judgments were associated with upstepped contours (downstep and upstep were defined as tonic syllables marked as beginning significantly lower or higher than the previous syllable, in a transcription based on Crystal, 1969). Thus prosodic effects may not necessarily be correctly interpreted by listeners, but some intonation patterns may nonetheless be reliably associated with continuations, others with finality. As this study also used dialogues which had been read (albeit by trained actors) and presented the utterances to listeners out of context, its results provide no direct evidence of the processing of turn structure in natural conversational situations.

Note finally that all these latter perceptual studies involve the simple judging of discourse structure; the studies of the incorporation of reference in a discourse model via probe and priming techniques have as yet no parallel in studies of topic or turn structure.

Summary

The studies reviewed in this section suggest that the prosodic realization of words is of direct relevance to the processing of discourse structure. Listeners appear to use prosodic information to predict upcoming locations of sentence accent. The evidence suggests that this is not a rhythmically-based prediction of stress locations but an active search for accented words because they are focussed. Furthermore, processing is facilitated by the placement of accent on new information, and the deaccenting of old information. A wide range of findings is consistent with the view that the relevant processing involves necessary integration of old concepts in an already existing discourse model; but the evidence so far does not allow us to decide whether the facilitation is evidence of direct exploitation of prosodic information in discourse-structure decisions, or arises indirectly via reference to the existing discourse model in the course of decoding poorer acoustic information. Only the study by Sedivy et al. (1995) has so far addressed on-line integration of prosodic and discourse structure. The suggestion by Terken and Nooteboom (1987) that the presence versus absence of accents would trigger bottom-up versus top-down processing respectively seems to us unwarranted; but there is clearly need for additional research. However, some of the effects that we have reviewed show potentially informative patterns. For instance, as pointed out by Donselaar (1995b), the relation of accentuation to new versus given information is asymmetrical. On the one hand, speakers seldom deaccent new information and if they do, this hinders listeners; on the other hand, they frequently accent given information and this interferes less with comprehension (Nooteboom & Terken, 1982; Terken & Nooteboom, 1987; Nooteboom & Kruyt, 1987). This suggests that prosodic information is perhaps not providing the listener with discourse-structure information in a direct manner. Accented

information is simply easier to process because the signal quality is higher; interference with processing by lowering the input quality harms comprehension more if the concepts involved are new rather than already available in the discourse model.

Similarities to the processing of accent for focus structure arise in the processing of the referential structure of discourse. Accent can mark concepts which will be referred to later, and can signal implicit reference. Furthermore, there is extensive evidence—though unsupported by any on-line processing data—that listeners can interpret prosodic information to derive cues to topic and turn structure in discourse.

The research that we have reviewed in this part has pointed up a number of limitations in discourse research. In most of the experiments on given and new information, "given" information was operationally defined in terms of repeated surface forms. The degree of givenness, and explicitness of anaphoric referring expressions can vary considerably (Garrod, Freudenthal, & Boyle, 1994). Production studies on the topic structure of discourse have not been adequately accompanied by studies of the perception of topic structure, and the perceptual studies that do exist did not use on-line paradigms. Natural spontaneous discourse has hardly been studied at all. In short, there are still many unanswered questions regarding the precise ways in which prosodic information is exploited in the processing of discourse structure.

CONCLUSION

In each of the areas of processing studies which we have reviewed, research involving prosody has focussed on one or two specific questions. In no case have the questions been cleared out of the way by definitive answers. However, as the summaries for each section endeavored to establish, progress has in each case been made towards a clearer picture of the role of prosody.

In the recognition of spoken words, the central issue has been whether or not prosodic information participates in the access code, that is, in the initial activation of word candidates. Models of spoken-word recognition distinguish stages in lexical processing: Initial access of lexical representations is often considered to be separable from the process of selection between potential candidate words activated by the input, and the latter process, in current computational models, ensues via a process of competition. In some models, the selection process in turn is separate from the processes of integration of the selected word with the rest of the utterance. This is not the place to review the models in question; it is sufficient to say that they differ principally in the extent to which these stages are held to be functionally separate or to overlap and /or interact.

From this point of view, it is remarkable that studies of prosody's role in word recognition have hardly contributed to distinguishing between individual models or classes of models. Moreover, although current models of the word-recognition process are in general computationally explicit, we know of no attempt to implement modeling of the prosodic structure of the input.[3] This is, in fact, one of the many ways in which the currently

[3] Most models operate in any case with fairly unrealistic input, defined for example as strings of separate phonemes. The absence of an explicit representation of prosodic structure is not the only way in which the nature of speech is inadequately captured by such an input form.

leading models reveal their origins in the modeling of word recognition in the visual modality.

Against the day when models are refined to take account of the prosodic structure of the input, our review indicates some general principles. The evidence suggests that prosodic information will participate in the initial activation stage of word recognition to the degree that its use is efficient. If the point at which prosodic information becomes useful is too late (because the word candidates it might have contributed to activating are already activated), or if it contributes no added value to the activation code over and above that which is contributed by segmental information, then the word-recognition process will not benefit from it. Thus, for example, although lexical prosody which is realized in a monosyllabic domain may be more likely to contribute in the initial activation of word candidates than lexical prosody which is realized in a polysyllabic domain, this is less a consequence of the domain size per se than an indirect effect of the late arrival of usable prosodic distinctions in the latter case.

In syntactic processing, the main questions with respect to prosody's role have been: does prosody serve to divide the input into major syntactically motivated chunks? Does prosodic information serve to resolve ambiguity, such that sentences which admit of more than one interpretation when they are written are effectively unambiguous when spoken? And is prosodic information consulted "on-line" in order to select between alternative syntactic analyses which present themselves, albeit temporarily, during the processing even of an unambiguous sentence?

The evidence at this processing level speaks against any direct availability of syntactic information in prosodic structure; prosodic hierarchies encode prosodic and not syntactic relationships. Furthermore, a given utterance may allow alternative prosodic structures which are equally likely and equally acceptable. It is only recently that studies of the role of prosody in syntactic analysis have accepted this to be the case, and have begun to approach the issue without the hypothesis of a one-to-one mapping.

Just as in the lexical arena, in theoretical treatments of syntactic processing, prosody has hardly been considered. The central distinctions between models of syntactic analysis again involve the autonomy versus interdependence of stages of processing. However, parts of the research on the role of prosody have in this case contributed to debate on the principal issues. Thus underlying many of the studies of the Minimal Attachment issue has been a general concern that the enormous body of research on this question, nearly all of it conducted in the visual modality, has been mis-targeted since there may in the spoken modality be no question about whether a verb-NP transition should be interpreted as the beginning of a direct object or of a clause complement. The contribution of prosodic studies to the theoretical debate has in this instance not been provision of further evidence regarding parsing preferences one way or the other, but consideration of the possibility that some parsing-preference questions may constitute a nonissue.

In the area of discourse-structure processing, the central issue has concerned whether prosodic realizations directly cue discourse relationships. Although the research evidence clearly suggests that prosodic information—in particular the accentuation or deaccentuation of lexical items—contributes to the construction of a discourse model, a directive role of prosody in this capacity does not seem to be indicated. Instead, all the results are consistent

with indirect explanations involving differences in the type of processing required by acoustically clearer versus less clear input.

The empirical evidence in this area is far less profuse than in the lexical and syntactic fields; but of course this is equally true of research in the three fields in general, irrespective of whether or not prosodic issues are addressed. Most of the studies that we reviewed were not directly motivated by the issues which have dominated theoretical models of discourse processing, however, and certainly do not involve attempts at explicit (computational) modeling.

The phrase "around the edge of language" has been applied to the role of prosody in linguistic structure (Bolinger, 1964). Does our review motivate the conclusion that studies of the role of prosody in spoken-language processing are around the edge of the current theoretical debates in this area of Psycholinguistics? To some extent this is true. However, we do not feel that this arises from the nature of the topic. Rather, to return to our opening, it is because prosody intrinsically belongs to spoken-language processing; it is the study of spoken-language processing itself which has lagged behind other areas in Psycholinguistics, and insofar as theories of spoken-language processing are grounded in earlier theories which were not specifically designed for that domain, they contain no ready role for prosody. It is up to the next generation of theoreticians of spoken-language processing to remedy this situation.

The next generation should also, we hope, engage in explicit debate about the definition of prosodic structure in relation to processing. Although the fact that prosody is largely ignored in many processing models has essentially arisen via a historical accident in Psycholinguistics, it has, at least in part, been responsible for the lack of unanimity to which we drew attention in the introduction: if prosody is not explicitly included in the model, there is little need to define it. The situation can then arise that researchers using the same kinds of methodology and the same terminology are in fact operating with different sets of assumptions about the object of their study. Note that we have observed this problem in reviewing the psycholinguistic literature; Nooteboom (1997) similarly calls attention to it in a review of phonetic studies.

Differing underlying assumptions may seem to plague the various sub-fields of prosodic studies within Psycholinguistics to differing degrees. In linguistics, the degree to which assumptions about prosody differ across theoretical perspectives is likely to be relatively clear. But as we pointed out in the introduction, studies addressing processing issues must base their research questions on the processing task: conversion of an input into a processed representation. This often has as a consequence that processing studies are agnostic as to linguistic theory, since if they base their empirical formulations on a particular theory, where alternative formulations based on alternative theoretical perspectives were possible, they may end up with a valuable addition to knowledge about the viability of the chosen theory but no definitive answer to the processing question. In the field of lexical processing, current models are comparatively explicit; as a result of this, researchers have paid close attention to the nature of the speech signal and in particular its temporal characteristics: what type of information is available when? This has led in turn to claims about the role of prosody in lexical processing being, in essence, deconstructed: if stress, for example, participates in activation of lexical entries, what features of the speech input correspond to stress information? How are such features processed, and when are they available to the processor?

In the areas of syntactic and discourse processing, modeling is certainly more computationally explicit now than it was a decade or two ago, but because of the long history of visual experiments in both these fields and the very short history of spoken-language work, models are, as yet, not explicit with respect to the characteristics of speech. At present these two sub-fields seem less diverse than the field of lexical processing in their assumptions about prosodic structure; but it is at least possible that this apparent unity will disappear as more computationally explicit models of processing — specifically in the speech domain—force the formulation of different questions.

Finally, the sheer length of this review attests to the fact that there is now a quite substantial body of work on the role of prosodic processing in comprehension. Nevertheless it will be obvious, especially from consideration of the summaries of each part, that our experience in considering the literature as a whole has been that it is very uneven. Some topics (the role of lexical stress in lexical access; the role of prosody in distinguishing between direct-object and clause-complement attachment; the signaling of given vs. new information) have been worried to pieces; others, even when they might have addressed exactly the same processing issues, have been ignored. That is, quite apart from issues of the theoretical framework within which research has been conceived, there has been a narrow choice among the possible topics such that whole areas remain to be discovered. We have attempted to highlight some obvious gaps in the literature, as well as to make clear what we consider to be the most productive current approaches, and we fervently hope that one function of this review will be to stimulate yet more research in this area, in particular in as yet unexplored directions. We hope to be out of date very soon!

Received: December 4, 1996; revised manuscript received: April 3, 1997; accepted May 10, 1997.

REFERENCES

ABERCROMBIE, D. (1967). *Elements of general phonetics*. Edinburgh, U.K.: Edinburgh University Press.

ABRAMS, K., & BEVER, T. G. (1969). Syntactic structure modifies attention during speech perception and recognition. *The Quarterly Journal of Experimental Psychology*, 21, 280-290.

ALLBRITTON, D. W., McKOON, G., & RATCLIFF, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memoir, and Cognition*, 22, 714-735.

ALTMANN, G. T. M., & CARTER, D. M. (1989). Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? *Computer Speech and Language*, 3, 265 - 275.

AULL, A. M. (1984). *Lexical stress and its application in large vocabulary speech recognition*. Unpublished master's thesis, M.I.T., Massachusetts.

AVESANI, C, HIRSCHBERG, J., & PRIETO, P. (1995). The intonational disambiguation of potentially ambiguous utterances in English, Italian, and Spanish. *Proceedings of the Thirteenth International Congress of Phonetic Sciences* (pp. 174- 177). Stockholm.

BANSAL, R. K. (1966). *The intelligibility of Indian English*. Unpublished doctoral dissertation, University of London, U.K.

BARD, E. G., COOPER, L., KOWTKO, J., & BREW, C. (1991). Psycholinguists studies on incremental recognition of speech: A revised and extended introduction to the messy and the sticky. *DYANA Deliverable* Rl .3B, University of Edinburgh.

BARD, E. G., LOWE, A. J, & ALTMANN, G. T. M. (1989). The effect of repetition on words in recorded dictations. In J. R Tubach & J. J. Mariani (Eds.), *Proceedings of the European Conference on Speech Communication and Technology* (pp. 573-576). Edinburgh.

BARD, E. G., SOTILLO, C, ANDERSON, A. H., DOHERTY-SNEDDON, G., & NEWLANDS, A. (1995). The control of intelligibility in running speech. *Proceedings of the Thirteenth International Congress of Phonetic Sciences* (pp. 188-191). Stockholm.

BARIK, H. C. (1977). Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech,* 20, 116-126.

BEACH, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language,* 30, 644-663.

BECKMAN, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes,* 11, 17-67.

BEVER, T. G,, LACKNER, J. R., & KIRK, R. (1969). The underlying structures of sentences are the primary units of intermediate speech processing. *Perception & Psychophysics,* 5, 225-234.

BEVER, T G., & MCELREE, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry,* 19, 35-43.

BIRCH, S., & CLIFTON, C. E. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech,* 38, 365-391.

BIRCH, S. L., & GARNSEY, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language,* 34, 232-267.

BLAAUW, E. (1994). The contribution of prosodic boundary markers to the difference between read and spontaneous speech. *Speech Communication,* 14,359-375.

BLUTNER, R., & SOMMER, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language,* 27, 359 - 367.

BOCK, J. K., & MAZZELLA, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory and Cognition,* 11, 64-76.

BOLINGER, D. L. (1964). Intonation: Around the edge of language. *Harvard Educational Review,* 34, 282-296.

BOLINGER, D. L. (1972). Accent is predictable (if you're a mindreader). *Language,* 48, 633-644.

BOLINGER, D. L. (1978). Intonation across languages. In J. H. Greenberg (Ed.), *Universals of human language, Vol. 2. Phonology* (pp. 471-524). Stanford: Stanford University Press.

BOND, Z. S. (1981). Listening to elliptic speech: Pay attention to stressed vowels. *Journal of Phonetics,* 9, 89-96.

BOND, Z. S., & G ARNES, S. (1980). Misperceptions of fluent speech. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 115 -132). Hillsdale, NJ: Erlbaum.

BOND, Z. S., & SMALL, L. H. (1983). Voicing, vowel and stress mispronunciations in continuous speech. *Perception & Psychophysics,* 34, 470-474.

BOUWHUIS, D., & de ROOIJ, J. J. (1977). Vowel length and the perception of prosodic boundaries. *IPO Annual Progress Report,* 12, 63-68.

BROWMAN, C. P. (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics,* 42.

BROWN, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler, & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 67-77). Heidelberg: Springer-Verlag.

BROWN, G., CURRIE, K,, & KENWORTHY, J. (1980). *Questions of intonation.* London: Croom Helm.

BRUBAKER, R. S. (1972). Rate and pause characteristics of oral reading. *Journal of Psycholinguistic Research,* 1, 141-147.

BRUCE, G. (1982). Textual aspects of prosody in Swedish. *Phonetica,* 39, 274-287.

BRUCE, G., GRANSTROM, B., GUSTAFSON, K., & HOUSE, D. (1992). Aspects of prosodic phrasing in Swedish. *Second International Conference on Spoken Language Processing* (pp. 109-112). Banff, Canada.

BURNHAM, D., FRANCIS, E,, WEBSTER, D., LUKSANEEYANAWIN, S., ATTAPAIBOON, C, LACERDA, E, & KELLER, P. (1996). Perception of lexical tone across languages: Evidence for

a linguistic mode of processing. *Proceedings of the Fourth International Conference on Spoken Language Processing (pp.* 2514-2516). Philadelphia.

BURNHAM, D., KIRKWOOD, K., LUKSANEEYANAWIN, S., & PANSOTTEE, S. (1992). Perception of Central Thai tones and segments by Thai and Australian adults. *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium of Language and Linguistics* (pp. 546-560). Bangkok: Chulalongkorn University Press.

BUTTERWORTH, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research,* 4,57-87.

BUXTON, H. (1983). Temporal predictability in the perception of English speech. In A. Cutler, & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 111-121). Heidelberg: Springer-Verlag.

CARLSON, R., GRANSTROM, B., LINDBLOM, B., & RAPP, K. (1972). Some timing and fundamental frequency characteristics of Swedish sentences: Data, rules, and a perceptual evaluation. Speech Transmission Laboratory (Stockholm): *Quarterly Progress and Status Report,* 4, 11 - 19.

CARTER, D. M. (1987). An information-theoretic analysis of phonetic dictionary access. *Computer Speech & Language,* 2, 1-11.

CASSIDY, K. W., & KELLY, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language,* 30, 348 - 369.

CHAFE, W. (1974). Language and consciousness. *Language,* 50, 111 - 133.

CHEN, H.-C, & CUTLER, A. (in press). Auditory priming in spoken and printed word recognition. In H.-C. Chen (Ed.), *The Cognitive Processing of Chinese and Related Asian Languages.* Hong Kong: Chinese University Press.

CHING, Y. C. (1985). Lipreading Cantonese with voice pitch. Paper presented to the 109th meeting, Acoustical Society of America, Austin (Abstract *Journal of the Acoustical Society of America,* 77, Supplement 1,S39-40).

CHING, Y C. (1988). Voice pitch information for the deaf. *Proceedings of the First Asian-Pacific Regional Conference on Deafness* (pp. 340-343). Hong Kong.

CLARK, H. H., & HAVILAND, S. E. (1977). Comprehension and the Given-New contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1 -40). Norwood, NJ: Ablex.

CLOITRE, M., & BEVER, T. G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes,* 3, 293 - 322.

COLE, R. A., & JAKIMIK, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America,* 67, 965 - 970.

COLE, R. A., JAKIMIK, I, & COOPER, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America,* 64,44-56.

COLLIER, R., & 't HART, J. (1975). The role of intonation in speech perception. In A. Cohen, & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 107-121). Heidelberg: Springer-Verlag.

COLLIER, R., de PIJPER, J. R., & SANDERMAN, A. A. (1993). Perceived prosodic boundaries and their phonetic correlates. *Proceedings of the DARPA Workshop on Speech and Natural Language* pp. 341 - 345). Princeton, NJ, March 21 - 24.

CONNINE, C. M., CLIFTON, C. E., & CUTLER, A. (1987). Lexical stress effects on phonetic categorization. *Phonetica,* 44, 133-146.

COOPER, W. E. (1976). The syntactic control of timing in speech production: A study of complement clauses. *Journal of Phonetics,* 4, 151-171.

COOPER, W. E., & PACCIA-COOPER, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

COOPER, W. E., & SORENSEN, J. M. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America,* 62, 682 - 692.

COOPER, W. E., & SORENSEN, J. M. (1981). *Fundamental frequency in sentence production.* New York: Springer-Verlag.

CROFT, W. (1995). Intonation units and grammatical structure. *Linguistics,* 33, 839-882.

CRYSTAL, D., & DAVY, D. (1969). *Investigating English style.* London: Longman.

CRYSTAL, T. H., & HOUSE, A. S. (1988). Segmental durations in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America,* 83, 1574-1585.

CRYSTAL, T. H., & HOUSE, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America,* 88, 101-112.

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

CUTLER, A. (1982). Prosody and sentence perception in English. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation* (pp. 201-216). London: LEA.

CUTLER, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech,* 29, 201-220.

CUTLER, A. (1987). Components of prosodie effects in speech recognition. *Proceedings of the Eleventh International Congress of Phonetic Sciences* (pp. 84-87). Tallinn, Estonia.

CUTLER, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech, and brain* (pp. 157-166). London: Macmillan.

CUTLER, A. (1995). Spoken word recognition and production. In J. L. Miller, & P. D. Eimas (Eds.), *Speech, language and communication* (pp. 97-136). [Vol. 11 of E. C. Carterette, & M. P. Friedman (Eds.), *Handbook of Perception and Cognition].* NY: Academic Press.

CUTLER, A., & BUTTERFIELD, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language,* 31, 218-236.

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

CUTLER, A., & CHEN, H.-C. (1995). Phonological similarity effects in Cantonese word recognition. *Proceedings of the Thirteenth International Congress of Phonetic Sciences* (pp. 106-109). Stockholm.

CUTLER, A., & CHEN, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics,* 59, 165-179.

CUTLER, A., & CLIFTON, C. E. (1984). The use of prosodie information in word recognition. In H. Bouma, & D. G. Bouwhuis (Eds.)., *Attention and performance X: Control of language processes* (pp. 183-196). Hillsdale, N.J.: Erlbaum.

CUTLER, A., & DARWIN, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics,* 29,217-224.

CUTLER, A., & FODOR, J. A. (1979). Semantic focus and sentence comprehension. *Cognition,* 7, 49-59.

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

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

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

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

CUTLER, A., & OTAKE, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language,* 33, 824 - 844.

CUTLER, A., & OTAKE, T. (1996). The processing of word prosody in Japanese. *Proceedings of the Sixth Australian International Conference on Speech Science and Technology* (pp. 599-604). Adelaide.

CUTLER, A., & PEARSON, M. (1986). On the analysis of prosodie turn-taking cues. In C. Johns-Lewis (Ed.), *Intonation in Discourse* (pp. 139-155). London: Croom Helm.

DAHAN, D. (1994). *Etude de la Prosodie du Francois en Parole Continue.* Unpublished doctoral dissertation, Universite Paris V, Paris.

DAHAN, D. (1996). The role of rhythmic groups in the segmentation of continuous French speech.

Proceedings of the Fourth International Conference on Spoken Language Processing (pp. 1185 -1188). Philadelphia.

DAHAN, D., & BERNARD, J. M. (1996). Interspeaker variability in emphatic accent production in French. *Language and Speech,* 39, 341-374.

DARWIN, C. J. (1975). On the dynamic use of prosody in speech perception. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 178-193). Berlin: Springer-Verlag.

DONSELAAR, W. van (1991). The function of prosody in speech perception. *Proceedings of the Twelfth International Congress of Phonetic Sciences* (pp. 466-469). Aix-en-Provence, Universite de Provence.

DONSELAAR, W. van (1995a). Listeners' use of the "information-accentuation" interdependence in processing implicit and explicit references. *Proceedings of the Fourth European Conference on Speech Communication and Technology* (pp. 979-982). Madrid.

DONSELAAR, W. van (1995b). *Effects of accentuation and given/new information on word processing.* Unpublished doctoral dissertation, University of Utrecht, The Netherlands.

DONSELAAR, W. van, KOSTER, M., & CUTLER, A. (in preparation). *Voornaam* is not a homophone: Lexical prosody and lexical access in Dutch.

DONSELAAR, W. van, & LENTZ, J. (1994). The function of sentence accents and given/new information in speech processing: Different strategies for normal-hearing and hearing-impaired listeners? *Language and Speech,* 37, 375-391.

DONZEL, M. E. van, & KOOPMANS-van BEINUM, F. J. (1996). Pausing strategies in discourse in Dutch. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1029 -1032). Philadelphia.

DUPOUX, E., & MEHLER, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language,* 29, 316-335.

DUPOUX, E., PALLIER, C, SEBASTIAN, N., & MEHLER, J. (1997). A destressing deafness in French? *Journal of Memory and Language,* 36,406 - 421.

EADY, S. J., & COOPER, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America,* 80, 402-415.

EEFTING, W. (1991). The effect of "information value" and "accentuation" on the duration of Dutch words, syllables and segments. *Journal of the Acoustical Society of America,* 89,412-424.

EPSTEIN, W. (1961). The influence of syntactical structure on learning. *American Journal of Psychology,* 74,80-85.

FEAR, B. D., CUTLER, A., & BUTTERFIELD, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America,* 97, 1893 -1904.

FELLOWES, J. M., REMEZ, R. E., & RUBIN, P E. (in press). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics.*

FERREIRA, F. (1993). Creation of prosody during sentence production. *Psychological Review.* 100, 233-253.

FERREIRA. F, ANES, M. D., & HORTNE, M. D. (1996). Exploring the use of prosody during language comprehension using the auditory moving window technique. *Journal ofPsycholinguistic Research,* 25,273-290.

FODOR, J. A., & BEVER, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior,* 4,414-420.

FOWLER, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech,* 31, 307-319.

FOWLER, C. A., & HOUSUM, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language,* 26,489 - 504.

FOX, R. A., & UNKEFER, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics,* 13, 69 - 90.

FRAZIER, L. (1978). *On comprehending sentences: Syntactic parsing strategies.* Bloomington: Indiana University Linguistics Club.

FRAZIER, L., & FODOR, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition,* 6,291-325.

FRAZ1ER, L., & RAYNER, K. (1987). Resolution of syntactic category ambiguities. *Journal of Memory and Language*, 26, 505-526.

GANDOUR, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, 9, 20-36.

GANDOUR, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149-175.

G ANONG, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.

GARRETT, M. (1964). *Structure and sequence in judgments of auditory events*. Unpublished doctoral dissertation, University of Illinois.

GARRETT, M., BEVER, X, & FODOR, J. (1965). The active use of grammar in speech perception. *Perception & Psychophysics*, 1, 30-32.

GARRO, L., & PARKER, F. (1982). Some suprasegmental characteristics of relative clauses in English. *Journal of Phonetics*, 10, 149-161.

GARROD, S., FREUDENTHAL, D., & BOYLE, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language*, 33, 39-68.

GEE, J. P., & GROSJEAN, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411 -458.

GEERS, A. E. (1978). Intonation contour and syntactic structure as predictors of apparent segmentation. *Journal of the Acoustical Society of America*, 4, 273-283.

GERKEN, L. A. (1994). Young children's representation of prosodic structure: Evidence from English-speakers' weak syllable omissions. *Journal of Memory and Language*, 33, 19-38.

GERKEN, L. A. (1996). Phonological and distributional cues to syntax acquisition. In J. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp.411 -425). Mahwah, NJ: Lawrence Erlbaum Ass.

GERNSBACHER, M. A., & JESCHENIAK, J. D. (1995). Cataphoric devices in spoken discourse. *Cognitive Psychology*, 29, 24-58.

GOLDMAN-EISLER, F. (1968). *Psycholinguistics: Experiment in spontaneous speech.* London: Academic Press.

GRABE, E., & WARREN, R (1995). Stress shift: Do speakers do it or do listeners hear it? In B. Conner! & A. Arvaniti (Eds.), *Papers in Laboratory Phonology IV* Cambridge, U.K.: Cambridge University Press.

GRABE, E., WARREN, P., & NOLAN, F. (1994). Resolving category ambiguities—evidence from stress shift. *Speech Communication*, 15, 101-114.

GRABE, E., WARREN, P., & NOLAN, F. (1995). Prosodic disambiguation of coordination structures. Paper presented to *Eighth Annual CUNY Conference on Human Sentence Processing.* Tucson, Arizona, March 16-18.

GROSJEAN, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics*, 21, 501 -529.

GROSJEAN, F, & GEE, J. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25,135 -155.

GROSJEAN, F., & HIRT, C. (1996). Using prosody to predict the end of sentences in English and French: Normal and brain-damaged subjects. *Language and Cognitive Processes*, 11, 107-134.

GROSZ, B., & HIRSCHBERG, J. (1992). Some intonational characteristics of discourse structure. *Proceedings of the Second International Conference on Spoken Language Processing* (pp. 429 - 432). Banff, Canada.

GROSZ, B., HIRSCHBERG, J., & NAKATANI, C. H. (1994). A study of intonation and discourse structure in directions. *Working papers of the Workshop on the Integration of Natural Language and Speech Processing, American Association for Artificial Intelligence*, 124-131.

GROSZ, B. J., & SIDNER, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12, 175-204.

GUSSENHOVEN, C. (1991). The English rhythm rule as an accent deletion rule. *Phonology*, 8,1-35.

HALLIDAY, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, 3, 199-244.

HARRIS, M. S., & UMEDA, N. (1974). Effect of speaking mode on temporal factors in speech: Vowel duration. *Journal of the Acoustical Society of America*, 56, 1016-1018.

't HART, J., COLLIER, R., & COHEN, A. (1990). *A perceptual study of intonation*. Cambridge, U.K.: Cambridge University Press.

HAWKINS, S., & WARREN, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*, 22,493-511.

HENDERSON, A., GOLDMAN-EISLER, R, & SKARBEK, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech*, 9, 207-216.

HEUVEN, V J. van (1985). Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position. *Journal of the Acoustical Society of America*, 78, s21.

HEUVEN, V J.van (1988). Effects of stress and accent on the human recognition of word fragments in spoken context: Gating and shadowing. *Proceedings of Speech '88, 7th FASE symposium* (pp. 811-818). Edinburgh.

HEUVEN, V J. van, & HAGMAN, P. J. (1988). Lexical statistics and spoken word recognition in Dutch. In P. Coopmans & A. Hulk (Eds.), *Linguistics in the Netherlands 1988* (pp. 59-68). Dordrecht: Foris.

HIRSCHBERG, I. (1993). Studies of intonation and discourse. *Working papers 41, ESCA workshop on prosody* (pp. 90-95). Lund.

HIRSCHBERG, J., & NAKATANI, C. H., (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the Thirty-fourth Annual Meeting of the Association for Computational Linguistics* (pp. 286-293). Santa Cruz.

HIRSCHBERG, I, & PIERREHUMBERT, J. (1986). The intonational structuring of discourse. *Proceedings of the Twenty-fourth Annual Meeting of the Association for Computational Linguistics* (pp. 134-144).

HIRSCHBERG, J., & WARD, G. (1991). Accent and bound anaphora. *Cognitive Linguistics*, 2-2,101-121.

HIRST, D. (1993). Detaching intonational phrases from syntactic structure. *Linguistic Inquiry*, 24,781 - 788.

HOUSE, D. (1990). *Tonal perception in speech*. Lund: Lund University Press.

HUNNICUTT, S. (1985). Intelligibility vs. redundancy — conditions of dependency. *Language and Speech*, 28, 47-56.

HUTTENLOCHER, D. P., & ZUE, V W. (1983). Phonotactic and lexical constraints in speech recognition. *Proceedings of the American Association for Artificial Intelligence* (pp. 172-176).

INKELAS, S., & ZEC, D. (1990). *The phonology-syntax connection*. Chicago: The University of Chicago Press.

JONGENBURGER, W (1996). *The role of lexical stress during spoken-word processing*. Unpublished doctoral dissertation, Leiden University, The Netherlands.

JONGENBURGER, W, & van HEUVEN, V J. (1995a). The role of linguistic stress in the time course of word recognition in stress-accent languages. *Proceedings of the Fourth European Conference on Speech Communication and Technology* (pp. 1695-1698). Madrid.

JONGENBURGER, W, & van HEUVEN, V J. (1995b). The role of lexical stress in the recognition of spoken words: prelexical or postlexical?, *Proceedings of the Thirteenth International Congress oj Phonetic Sciences* (pp. 368-371). Stockholm.

JUN, S. A., & OH, M. (1996). A prosodic analysis of three types of wh-phrases in Korean. *Language and Speech*, 39, 37-61.

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

KELLY, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99, 349-364.

KELLY, M. H., & BOCK, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 389-403.

KENNEDY, A., MURRAY, W. S., JENNINGS, E, & RE1D, C. (1989). Parsing complements: Comments on the generality of the principle of minimal attachment. *Language and Cognitive Processes, 4*, 51-76.

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

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

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

KOLINSKY, R., MORAIS, J., & CLUYTENS, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language,* 34, 19-40.

KONG, Q. M. (1987). Influence of tones upon vowel duration in Cantonese. *Language and Speech,* 30, 387-399.

KOOPMANS-van BEINUM, F. J., & BERGEM, D. R. van (1989). The role of "given" and "new" in the production and perception of vowel contrasts in read text and in spontaneous speech. *Proceedings of the European Conference on Speech Communication and Technology* (pp. 113-116). Edinburgh.

KOSTER, M., & CUTLER, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in Dutch. *Proceedings of the Fifth European Conference on Speech Communication and Technology* (2167 - 2170). Rhodes.

KOWAL, S., O'CONNELL, D,, O'BRIEN, E. A., & BRYANT, E. T. (1975). Temporal aspects of reading aloud and speaking: Three experiments. *American Journal of Psychology,* 88, 549-569.

KRATOCHVTL, P. (1971). An experiment in the perception of Peking dialect tones. In I.-L. Hansson (Ed.), *A Symposium on Chinese Grammar* (pp. 7-31). Lund: Curzon Press.

KREIMAN, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics,* 10, 163-175.

KUTIK, E. J., COOPER, W. E., & BOYCE, S. (1983). Declination of fundamental frequency in speakers* production of parenthetical and main clauses. *Journal of the Acoustical Society of America,* 73, 1731-1738.

LADD, D. R. (1980). *The Structure of intonational meaning.* Bloomington: Indiana University Press.

LADD, D. R. (1988). Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America,* 84, 530-544.

LADD, D. R. (1996). *Intonational phonology.* Cambridge, U.K.: Cambridge University Press.

LADD, D. R., & CUTLER, A. (1983). Models and measurements in the study of prosody. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 1-10). Heidelberg: Springer-Verlag.

LAGERQUIST, L. M. (1980). Linguistic evidence from paranomasia. *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society,* 185-191.

LARKEY, L. S., & DANLY, M. (1983). Fundamental frequency and sentence comprehension. *MIT Speech Communication Group Working Papers, 2,* 25-39.

LEE, L., & NUSBAUM, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics,* 53, 157-165.

LEE, Y.-S., VAKOCH, D. A., & WURM, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal ofPsycholinguistic Research,* 25, 527-542.

LEHISTE, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

LEHISTE, I. (1972). Timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America,* 51, 2018-2024.

LEHISTE, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa,* 7, 107-122.

LEHISTE, I. (1979). Perception of sentence and paragraph boundaries, In B. Lindblom & S. Oehman (Eds.), *Frontiers of speech communication research* (pp. 191-201). New York: Academic Press.

LEHISTE, I. (1980). Phonetic characteristics of discourse. *Acoustical Society of Japan, Transactions of the Committee on Speech Research,* 25-38.

LEHISTE, I., OLIVE, J. P., & STREETER, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America,* 60, 1199 -1202.

LEHISTE, 1., & WANG, WS.-Y. (1977). Perception of sentence and paragraph boundaries with and without

semantic information. In W. U. Dressier & O. E. Pfeiffer (Eds.), *Phonologica 1976* (pp. 277-283). Innsbruck: Institut fur Sprachwissenschaft der Universitat Innsbruck.

LEONARD, L. B. (1974). The role of intonation in the recall of various linguistic stimuli. *Language and Speech*, 16, 327-335.

LEVIN, H., SCHAFFER, C. A., & SNOW, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Language and Speech*, 25,43 - 54.

LEYDEN, K. van, & HEUVEN, V J. van (1996). Lexical stress and spoken word recognition: Dutch vs. English. In C. Cremers & M. den Dikken (Eds.), *Linguistics in the Netherlands 1996* (pp. 159 - 170). Amsterdam: John Benjamins.

LIBERMAN, M., & PRINCE, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 2,249-336.

LIEBERMAN, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6, 172-187.

LIEBERMAN, P. (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.

LIN, H.-B., & REPP, B. H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, 32,25-44.

MacDONALD, M. C, PEARLMUTTER, N. J., & SEIDENBERG, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

MacLAY H., & OSGOOD, C.E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15, 19-44.

MARSLEN-WILSON, W. S., TYLER, L. K., WARREN, P., GRENIER, P., & LEE, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology*, 45A, 73-87.

MARTIN, J. G. (1968). Temporal word spacing and the perception of ordinary, anomalous, and scrambled strings. *Journal of Verbal Learning and Verbal Behaviour*, 7, 154-157.

MARTIN, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 79, 487-509.

MARTIN, J. G. (1979). Rhythmic and segmental perception are not independent. *Journal of the Acoustical Society of America*, 65, 1286-1297.

MATTYS, S. L., & SAMUEL, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language*, 36, 87-116.

MAZUKA, R., MISONO, Y, TADAHISA, K., & KIRITANI, S. (1997). *Levels of informativeness of prosodic cues for resolving syntactic ambiguity*. Paper presented at the 10th Annual CUNY Conference on Human Sentence Processing, Santa Monica, CA.

McALLISTER, J. (1991). The processing of lexically stressed syllables in read and spontaneous speech. *Language and Speech*, 34, 1 - 26.

McKOON, G., & RATCLIFF, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behaviour*, 19, 668 - 682.

McQUEEN, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 17,433-443.

McQUEEN, J. M., NORRIS, D G., & CUTLER, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 621 -638.

MEHLER, J., DOMMERGUES, J.-Y, FRAUENFELDER, U, & SEGUI, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20, 298-305.

MEHTA, G., & CUTLER, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech*, 31, 135 -156.

MELTZER, R. H., MARTIN, J. G., MILLS, C. B., IMHOFF, D. L., & ZOHAR, D. (1976). Reaction time to temporally displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 277-290.

MENN, L., & BOYCE, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, 25,341-383.

MENS, L., & POVEL, D. (1986). Evidence against a predictive role for rhythm in speech perception. *The Quarterly Journal of Experimental Psychology,* **38A,** 177-192.

MILLER, G. A., & ISARD, S. D. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behaviour,* 2,217 - 228.

MILLER, J. L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics,* 24, 175-180.

MISONO, Y., MAZUKA, R., KONDO, X, & KIRITANI, S. (1997). Effects and limitations of prosodic and semantic biases on syntactic disambiguation. *Journal of Psycholinguistic Research,* 26, 229-245.

MITCHELL, D. C. (1994). Sentence parsing. In M. A. Gerasbacher (Ed.), *Handbook of Psycholinguistics* (pp 375-409). San Diego, CA: Academic Press.

MONNIN, P., & GROSJEAN, F. (1993). Les structures de performance en frangais: caracterisation et prediction. *L'Annee Psychologique,* 93, 9-30.

MURRAY, W. S., WATT, S. M., & KENNEDY, A. (submitted). Parsing ambiguities: Modality, processing options, and the garden path.

NAGEL, H. N., SHAPIRO, L. P., & NAWY, R. (1994). Prosody and the processing of filler-gap sentences. *Journal of Psycholinguistic Research,* 23 (6), 473 -485.

NAKATANI, L. H., & SCHAFFER, J. A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America,* 63, 234-245.

NEEDHAM, W. P. (1990). Semantic structure, information structure, and intonation in discourse production. *Journal of Memory and Language,* 29, 455-468.

NESPOR, M., & VOGEL, I. (1983). Prosodic structure above the word. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models andmeasurements* (pp. 123-140). Heidelberg: Springer-Verlag.

NESPOR, M., & VOGEL, I. (1986). *Prosodic phonology.* Dordrecht: Foris.

NICOL, J. L. (1996). What can prosody tell a parser? *Journal of Psycholinguistic Research,* 25,179 -192.

NICOL, J. L., FODOR, J. D., & SWINNEY, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 20, 1229-1238.

NICOL, J. L., & PICKERING, M. J. (1993). Processing syntactically ambiguous sentences: Evidence from semantic priming. *Journal of Psycholinguistic Research,* 22, 207-237.

NISHINUMA, Y. (1994). How do the French perceive tonal accent in Japanese? Experimental evidence. *Proceedings of the Third International Conference on Spoken Language Processing* (pp. 1739-1742). Yokohama.

NISHINUMA, Y, ARAI, M., & AYUSAWA, T. (1996). Perception of tonal accent by Americans learning Japanese. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 646-649). Philadelphia.

NOOTEBOOM, S. G. (1995). Limited lookahead in speech production. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing speech: Contemporary issues—for Katherine Safford Harris* (pp. 1-18). Woodbury, NY: AIP Press.

NOOTEBOOM, S. G. (1997). The prosody of speech: Melody and rhythm. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 640-673). Oxford: Blackwell Publishers.

NOOTEBOOM, S. G., BROKX, J. P. L., & ROOIJ, J. J. de (1978). Contributions of prosody to speech perception. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 75 - 107). Chichester: John Wiley & Sons.

NOOTEBOOM, S. G., & KRUYT, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America,* 82, 1512-1524.

NOOTEBOOM, S. G,, & TERKEN, J. M. B. (1982). What makes speakers omit pitch-accents? An experiment. *Phonetica,* 39, 317-336.

O'CONNELL, D. C, TURNER, E. A., & ONUSKA, L. A. (1968). Intonation, grammatical structure, and contextual association in free recall. *Journal of Verbal Learning and Verbal Behaviour,* 7,110 - 116.

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

O'SHAUGHNESSY, D. (1979). Linguistic features in fundamental frequency patterns. *Journal of Phonetics,* 7, 119-145.

OTAKE, T, HATANO, G., CUTLER, A., & MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language,* 32, 358-378.

OTAKE, T, HATANO, G., & YONEYAMA, K. (1996). Speech segmentation by Japanese listeners. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies,* (pp. 183-201). Berlin: Mouton de Gruyter.

OTAKE, T, YONEYAMA, K., CUTLER, A., & van der LUGT, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America,* 100, 3831 - 3842.

PALLIER, C, SEBASTIAN-GALLES, N., FELGUERA,T, CHRISTOPHE, A., & MEHLER, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language,* 32,373-389.

PERETZ, I., LUSSIER, I., & BELAND, R. (1996). The roles of phonological and orthographic code in word stem completion. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 217-226). Berlin: Mouton de Gruyter.

PERROT, J. (1978). Fonctions syntaxiques, enonciation, information. *Bulletin de la Societe Linguistique de Paris,* 53, 85-101.

PIERREHUMBERT, J., & HIRSCHBERG, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271 -323). Cambridge, MA: MIT Press.

PIJPER, J. R. de, & SANDERMAN, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America,* 96,203 7 - 2047.

PIKE, K. L. (1945). *The intonation of American English.* Ann Arbor: University of Michigan Press.

PITT, M. A., & SAMUEL, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance,* 16, 564-573.

PORT, R. E, REILLY, W. X, & MAKI, D. P. (1988). Use of syllable-scale timing to discriminate words. *Journal of the Acoustical Society of America,* 83, 265-273.

PRICE, P., & OSTENDORF, M. (1996). Combining linguistic with statistical methods in modeling prosody. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 67-83). Mahwah, NJ: Lawrence Erlbaum Ass.

PRICE, P. J., OSTENDORF, M., SHATTUCK-HUFNAGEL, S., & FONG, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America,* 90, 2956-2970.

PRINCE, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-255). New York: Academic Press.

PYNTE, J., & PRIEUR, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes,* 11, 165-191.

QUENE, H. (1987). Perceptual relevance of acoustical word boundary markers. *Proceedings of the Eleventh International Congress of Phonetic Sciences* (pp. 79-82). Tallinn. Estonia.

QUENE, H. (1989). *The influence of acoustic-phonetic word boundary markers on perceived word segmentation in Dutch.* Unpublished doctoral dissertation, University of Utrecht, The Netherlands.

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

QUENE, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America,* 94, 2027-2035.

READ, C, KRAAK, A., & BOVES, L. (1980). The interpretation of ambiguous who-questions in Dutch: The effect of intonation. In W. Zonneveld & F. Weerman (Eds.), *Linguistics in the Netherlands* 1977-1979 (pp. 389-410). Dordrecht: Foris.

REMEZ, R. E., FELLOWES, J. M., & RUBIN, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance,* 23,651 - 666.

REMEZ, R. E., RUBIN, P. E., PISONI, D. B., & CARRELL, T. D. (1981). Speech perception without traditional speech cues. *Science,* 212, 947-950.

REPP, B. H., & LIN, H.-B. (1990). Integration of segmental and tonal information in speech perception. *Journal of Phonetics*, 18,481-495.

RICE, K. D. (1987). On defining the intonational phrase: evidence from Slave. *Phonology Yearbook*, 4, 37-59.

RIETVELD, A. C. M. (1980). Word boundaries in the French language. *Language and Speech*, 23, 289-296.

ROBINSON, G. M. (1977). Rhythmic organization in speech processing. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 83-91.

ROOIJ, J. J. de (1975). Prosody and the perception of syntactic boundaries. *IPO Annual Progress Report*, 10,36-39.

ROOIJ, J. J. de (1976). Perception of prosodic boundaries. *IPO Annual Progress Report*, 11, 20-24.

SAGISAKA, Y., CAMPBELL, N, & HIGUCHI, N. (1997). *Computing prosody: Computational models for processing spontaneous speech.* New York: Springer-Verlag.

SANDERMAN, A. (1996). *Prosodic phrasing: Production, perception, acceptability, and comprehension.* Unpublished doctoral dissertation, University of Eindhoven, The Netherlands.

SANTEN, J. P. H. van, & OLIVE, J. P. (1990). The analysis of contextual effects on segmental duration. *Computer Speech & Language*, 4, 359-390.

SCHAFER, A. (1995). The role of optional prosodic boundaries. *Paper presented to the Eighth Annual CUNY Conference on Human Sentence Processing.* Tucson, Arizona. March 16-18.

SCHAFER, A., CARTER, J., CLIFTON J. R. C, & FRAZIER, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, 11, 135-163.

SCHOLES, R. J. (1971). On the spoken disambiguation of superficially ambiguous sentences. *Language and Speech*, 14, 1-11.

SCHREUDER, R., & BAAYEN, R. H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, 33, 357-375.

SCOTT, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, 71, 996-1007.

SCOTT, D. R., & CUTLER, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, 23, 450 - 466.

SEDIVY, J., TANENHAUS, M., SPTVEY-KNOWLTON, M., EBERHARD, K., & CARLSON, G. (1995). Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye movements to a visual model. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 375 - 380). Hillsdale, NJ: Erlbaum.

SEGUI, J., FRAUENFELDER, U. H., & MEHLER, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology*, **72,**471-477.

SELKIRK, E. (1984). *Phonology and Syntax: the relation between sound and structure.* Cambridge, MA: MIT Press.

SELKIRK, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, 3, 371-405.

SHANNON, R. V, ZENG. F.-G., KAMATH, V, WYGONSKI, J., & EKELID, M. (1995). Speech recognition with primarily temporal cues. *Science*, **270,** 303-304.

SHATTUCK-HUFNAGEL, S., & TURK, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193-247.

SHEN, X. S., & LIN, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, 34, 145-156.

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.

SHILLCOCK, R., BARD, E. G., & SPENSLEY, F. (1988). Some prosodic effects on human word recognition in continuous speech. *Proceedings of Speech '88, 7th FASE Symposium* (pp. 827 - 834). Edinburgh.

SILVERMAN, K. E. A. (1987). *The structure and processing of fundamental frequency contours.* Unpublished doctoral dissertation, University of Cambridge, Cambridge, U.K.

SILVERMAN, K. E. A., KALYANSWAMY, A., SILVERMAN, J., BASSON, S., & YASHCHIN, D. (1993). Synthesiser intelligibility in the context of a name-and-address information service. *Proceedings of the Third European Conference on Speech Communication and Technology* (pp.2169-2172). Berlin.

SLOWIAZCEK, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech,* 33,47-68.

SLUIJTER, A., & TERKEN, J. M. B. (1993). Beyond sentence prosody: Paragraph intonation in Dutch. *Phonetica,* 50, 180-188.

SON, R. J. J. H. van & SANTEN, J. P. H. van (1997). Strong interaction between factors influencing consonant duration. *Proceedings of the Fifth European Conference on Speech Communication and Technology* (319 - 322). Rhodes.

SORIN, C. (1981). Functions, roles, and treatments of intensity in speech. *Journal of Phonetics,* 9, 359-374.

SORIN, C, & Le BRAS, J. (1983). Role of FO contours on sentence identification in noise. *Abstracts of the Tenth International Congress of Phonetic Sciences* (p. 587). Dordrecht: Foris.

SPEER, S. R., CROWDER, R. G., & THOMAS, L. M. (1993). Prosodic structure and sentence recognition. *Journal of Memory and Language,* 32, 336-358.

SPEER, S. R., KJELGAARD, M. M., & DOBROTH, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal ofPsycholinguistic Research,* 25,247-268.

SPEER, S. R, SHUT, C.-L., & SLOWIACZEK, M. L. (1989). Prosodic structure in language understanding: Evidence from tone sandhi in Mandarin. *Language and Speech,* 32, 337-354.

STEEDMAN, M. (1990). Syntax and intonational structure in a combinatory grammar. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 457-482). Cambridge, MA: MIT Press.

STEEDMAN, M. (1991). Structure and intonation. *Language,* 67, 260-296.

STIRLING, L., & WALES, R. (1996). Does prosody support or direct sentence processing? *Language and Cognitive Processes,* 11, 193-212.

STRANGE, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America,* 85, 2135-2153.

STRAUB, K. A. (1996). Prosodic cues in syntactically ambiguous strings: An interactive speech planning mechanism. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1640-1643). Philadelphia.

STREETER, L. A. (1978). Acoustic determinants of phrase boundary location. *Journal ofthe Acoustical Society of America,* 64, 1582-1592.

SUOMI, K., McQUEEN, J. M., & CUTLER, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language,* 36,422 - 444.

SVENSSON, S.G. (1971). A preliminary study of the role of prosodic parameters in speech perception. *Speech Transmission Laboratory (Stockholm): Quarterly Progress and Status Report,* 2-3,24-42.

SVENSSON, S.G. (1974). Prosody and grammar in speech perception. *Monographs from the Institute of Linguistics,* University of Stockholm, 2.

SWERTS, M. (1997). Prosodic features at discourse boundaries of different strengths. *Journal of the Acoustical Society of America,* 101, 514-521.

SWERTS, M., BOUWHUIS, D. G., & COLLIER, R. (1994). Melodic cues to the perceived "finality" of utterances. *Journal of the Acoustical Society of America,* 96, 2064-2075.

SWERTS, M., & GELUYKENS, R. (1993). The prosody of information units in spontaneous monologue. *Phonetica,* 50, 189-196.

SWERTS, M., & GELUYKENS, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech,* 37 (1), 21 - 43.

SWERTS, M., WICHMANN, A., & BEUN, R.-J. (1996). Filled pauses as markers of discourse structure. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1033- 1036). Philadelphia.

SWINNEY, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior,* 18, 645 - 659.

TAFT, L. (1984). *Prosodic constraints and lexical parsing strategies.* Unpublished doctoral dissertation, University of Massachusetts, Massachusetts.

TAFT, M., & CHEN, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H.-C. Chen, & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 151 -172). Amsterdam: Elsevier.

TERKEN, J., & HIRSCHBERG, J. (1994). Deaccentuation of words representing "given" information: Effects of persistence of grammatical function and surface position. *Language and Speech,* 37, 125-145.

TERKEN, J., & LEMEER, G. (1988). Effects of segmental quality and intonation on quality judgments for texts and utterances. *Journal of Phonetics,* 16, 453 -457.

TERKEN, J., & NOOTEBOOM, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2,* 145-163.

THORSEN, N. G. (1985). Intonation and text in Standard Danish. *Journal of the Acoustical Society of America,* 77, 1205-1216.

TSANG, K. K., & HOOSAIN, R. (1979). Segmental phonemes and tonal phonemes in comprehension of Cantonese. *Psychologia,* 22, 222-224.

TYLER, L. K., & WARREN, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language,* 26, 638-657.

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

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

VAISSIERE, J. (1974). On French prosody. *Quarterly Progress Report, M.I.T,* **114,** 212-223.

VAISSIERE, J. (1975). Further note on French prosody. *Quarterly Progress Report, M.I.T,* **115,**251-262.

VAISSIERE, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 53-66). Heidelberg: Springer-Verlag.

VALLDUVI, E. (1992). *The informational component.* New York: Garland Press.

VANCE, T J. (1987). *An introduction to Japanese phonology.* Albany: SUNY Press.

VENDITTI, J. J., JUN, S.-A., & BECKMAN, M. E. (1996). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 287-311). Hillsdale, NJ: Erlbaum.

VENDITTI, J. J, & SWERTS, M. (1996). Intonational cues to discourse structure in Japanese. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 725 - 728). Philadelphia.

VENDITTI, J. J., & YAMASHITA, H. (1994). Prosodic information and processing of temporarily ambiguous constructions in Japanese. *Proceedings of the International Conference of Speech and Language Processing* (pp. 1147-1150). Yokohama, Japan.

VROOMEN, J., ZON, M. van, & GELDER, B. van (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition,* 24, 744-755.

WAIBEL, A. (1988). *Prosody and speech recognition.* London: Pitman.

WALES, R., & TONER, H. (1979). Intonation and ambiguity. In W. E. Cooper, & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 135- 158). Hillsdale, N.J.: Erlbaum.

WALSH DICKEY, L. (1996). Limiting-domains in lexical access: Processing of lexical prosody. In M. Dickey & S. Tunstall (Eds.), *University of Massachusetts Occasional Papers in Linguistics, 19: Linguistics in the Laboratory* (pp. 133-155).

WARREN, P. (1985). *The temporal organization and perception of speech.* Unpublished doctoral dissertation, University of Cambridge, Cambridge, U.K.

WARREN, P., GRABE, E., & NOLAN, F. (1995). Prosody, phonology, and parsing in closure ambiguities. *Language and Cognitive Processes,* 10,457-486.

WATT, S. M., & MURRAY, W. S. (1996). Prosodic form and parsing commitments. *Journal of Psycholinguistic Research,* 25, 291-318.

WHALEN, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics,* 50, 351-360.

WIGHTMAN, C. W,, SHATTUCK-HUFNAGEL, S., OSTENDORF, M,, & PRICE, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America,n,\lQl-\l\l.*

WINGFIELD, A. (1975a). The intonation-syntax interaction: Prosodic features in perceptual processing of sentences. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 146-160). Berlin: Springer-Verlag.

WINGFIELD, A. (1975b). Acoustic redundancy and the perception of time-compressed speech. *Journal of Speech and Hearing Research,* 18, 96-104.

WINGFIELD, A., & KLEIN, J. F. (1971). Syntactic structure and acoustic pattern in speech perception. *Perception & Psychophysics,* 9, 23-25.

WINGFIELD, A., BUTTET J., & SANDOVAL, W. (1979). Intonation and intelligibility of time-compressed speech. Supplementary report: English vs. French. *Journal of Speech & Hearing Research,* 22, 708 -716.

YUEN, 1.(1995). *An exploratory study ofFO height in Cantonese tone perception.* Unpublished master's thesis, University of Edinburgh, Edinburgh, U.K.

ZURIF, E. B., & MENDELSOHN, M. (1972). Hemispheric specialization for the perception of speech sounds: The influence of intonation and structure. *Perception & Psychophysics,* 11, 329-332.