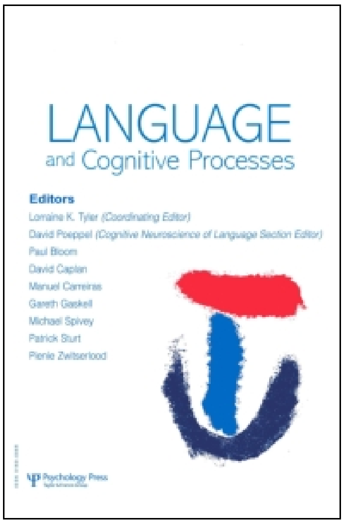

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683153

Models of continuous speech recognition and the contents of the vocabulary

James M. McQueen [a]; Anne Cutler [a]; Ted Briscoe [b]; Dennis Norris [c]
[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands [b] Computer Laboratory, University of Cambridge, Cambridge, UK [c] Medical Research Council Applied Psychology Unit, Cambridge, UK

Online Publication Date: 01 August 1995

To cite this Article McQueen, James M., Cutler, Anne, Briscoe, Ted and Norris, Dennis(1995)'Models of continuous speech recognition and the contents of the vocabulary',Language and Cognitive Processes,10:3,309 — 331

To link to this Article: DOI: 10.1080/01690969508407098

URL: http://dx.doi.org/10.1080/01690969508407098

LANGUAGE AND COGNITIVE PROCESSES, 1995, *10* (3/4), 309–331

Models of Continuous Speech Recognition and the Contents of the Vocabulary

James M. McQueen, Anne Cutler

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Ted Briscoe

Computer Laboratory, University of Cambridge, Cambridge, UK

Dennis Norris

Medical Research Council Applied Psychology Unit, Cambridge, UK

Several models of spoken word recognition postulate that recognition is achieved via a process of competition between lexical hypotheses. Competition not only provides a mechanism for isolated word recognition, it also assists in continuous speech recognition, since it offers a means of segmenting continuous input into individual words. We present statistics on the pattern of occurrence of words embedded in the polysyllabic words of the English vocabulary, showing that an overwhelming majority (84%) of polysyllables have shorter words embedded within them. Positional analyses show that these embeddings are most common at the onsets of the longer word. Although both phonological and syntactic constraints could rule out some embedded words, they do not remove the problem. Lexical competition provides a means of dealing with lexical embedding. It is also supported by a growing body of experimental evidence. We present results which indicate that competition operates both between word candidates that begin at the same point in the input and candidates that begin at different points (McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, in press). We conclude

Requests for reprints should be addressed to James M. McQueen, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands. E-mail: james@mpi.nl.

This work was supported by Grant No. E304/148 from the Joint Councils Initiative in Cognitive Science and HCI. We thank the Longman Group UK, Inc., for allowing us to use a machine-readable version of the *Longman Dictionary of Contemporary English*. We would like to acknowledge the contributions of Gerry Altmann, Lolly Tyler, Ellen Bard and an anonymous reviewer to this paper. This work was carried out when both J.M. and A.C. were at the MRC Applied Psychology Unit, Cambridge, UK.

that lexical competition is an essential component in models of continuous speech recognition.

INTRODUCTION

The effortless way in which we as listeners hear spoken language as a discontinuous sequence of individual words might lead us to expect that the physical speech stream provides us with this discontinuous sequence. But it does not. Spoken language is continuous, with few consistent and determinate cues to word boundaries (Lehiste, 1972; Nakatani & Dukes, 1977). Not only do we have to recognise words in an utterance, we also have to determine where the words begin in the utterance. Continuous speech recognition therefore entails an analysis of unsegmented input into a segmented string of words.

Two general strategies have been proposed to solve the segmentation problem. The first postulates strictly sequential processing, in which a single interpretation of a section of the input is settled upon prior to processing of the following section of input. On this account, words must be recognised in the order they were spoken. The second solution postulates delayed commitment, in which multiple interpretations of the input are considered in parallel, and a unique interpretation may be delayed until the arrival of subsequent disambiguating information. Several models of spoken word recognition have been proposed which are based on sequential recognition, and others have been proposed which involve delayed commitment. Although both approaches assume that speech is processed incrementally, strictly sequential models make the stronger claim that words uttered later in time cannot influence the recognition of those uttered earlier.

Early models of spoken word recognition were strictly sequential (Cole & Jakimik, 1978; 1980; Marslen-Wilson & Welsh, 1978). In these models, word boundaries emerge from the recognition of individual words. When the current word has been recognised, its offset will specify where the onset of the following word is. Words sharing the same initial portion are all accessed during recognition, but only those words whose onsets are aligned are ever considered at the same time. Recognition and segmentation in sequential models therefore proceed in strict order, exactly the order in which the words were spoken. The correct placement of a word boundary and the subsequent recognition of the following word depend on successful recognition of the preceding word, and hence on accurate placement of the preceding word boundary. If listeners hear *a Christmas pudding*, they will be able to recognise *Christmas* after they have heard /krɪsm/ (since Christmas is the only possible completion of this string). They will then be able to postulate a word boundary after the /əs/, and to start processing words (including *pudding*) beginning with the /p/. This, however, all depends on successful recognition of *a*, and placement of a boundary before the /k/, and this cannot be achieved until words like *acrylic* have been ruled out.

Strictly sequential models have been called into question by statistical analysis of vocabulary. Luce (1986) used lexical statistics to highlight the implausibility of these models. Their plausibility depends, in part, on the proportion of words which become unique (like *Christmas*) before their offset. Luce found that 41% of the words in a 20,000-word dictionary did not become unique before their final phoneme. Furthermore, many of the words which did become unique were long and of low frequency. In an analysis weighted by frequency, Luce showed that the probability of a word not becoming unique before its final phoneme was 0.61 (0.23 for words becoming unique on their final phoneme, and 0.38 for words becoming unique after their last phoneme). Many words, particularly short words, cannot be recognised before their offset, as would be required in a strictly sequential model.

There is also experimental evidence against sequential models. Listeners have been presented with incrementally longer and longer portions of spoken sentences and have been asked to identify the words they hear (the gating task) (Bard, Shillcock, & Altmann, 1988; Grosjean, 1985). It was found that many words, particularly short ones, could not be recognised before the onset of the following word had been heard. That is, listeners were unable to recognise words reliably before their offset, and hence could not predict where the words which followed them began. These results suggest that listeners use following context for word recognition, contrary to the claim of sequential recognition models that recognition proceeds word by word.

Strictly sequential processing has therefore not been proposed in more recent models of spoken word recognition. Models such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994a) postulate delayed commitment, in which following context can be used to influence the recognition of words spoken earlier. The Cohort model, though sequential in its first formulation (Marslen-Wilson & Welsh, 1978), has been aligned with the TRACE model (Marslen-Wilson, 1987), and thus also allows for delayed commitment. In the earliest formulation, only words consistent with a word onset became cohort members, and, as bottom-up information inconsistent with certain cohort members became available, those cohort members were ruled out, until only one word remained. This word could then be recognised, and, in keeping with sequential processing, the next word onset could be postulated, and only then could a new cohort be generated. In the revised model, as in TRACE and Shortlist, there is continuous activation of candidate words, wherever they might begin in the input. At any moment, multiple lexical hypotheses, spanning different portions of the input, will have been activated by the signal.

In all three models, evidence for candidate words is represented by differential degrees of activation. The degree of activation of different words is determined by the acoustic–phonetic match between each word and the signal. Acoustic–phonetic information available in a subsequent word (such as /ɪt/ in the phrase *can it*) can decrease the activation levels of competitors of *can*, such as *candid*, assisting the recognition of *can*. In this way, all three models delay commitment to *can* until after its offset, and thus are not strictly sequential.

In TRACE and Shortlist, recognition is based on direct competition between candidate words. Word candidates spanning the same portions of the input, partially overlapping portions, and even completely different portions, are all active simultaneously, and all compete together. Recognition of a word at time t is therefore partially determined by the evidence for words at times $t-1$, $t+1$, $t+2$, and so on. In both TRACE and Shortlist, there are direct inhibitory connections between different words. The architecture of TRACE is such that the number of inhibitory connections required for large vocabulary recognition is unrealistically large, making TRACE highly implausible (Norris, 1994a). In Shortlist, this problem is overcome by strict limitation on the number of words that can be activated at any one time. Only those words which best match the input (the "shortlist") are activated and allowed to enter into the competition process. The model operates successfully with a lexicon of over 25,000 words (McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, in press).

Lexical competition is a powerful mechanism for continuous speech recognition. It provides the major benefit of sequential processing (recognition of words as soon as there is sufficient evidence) without its major drawback (dependence on recognition of all words in strict order). If *Christmas pudding* is presented to a competition model, *Christmas* will win out early in the competition process, since *Christmas* becomes unique before its offset and other candidates will match the input much less well. The earliness of recognition will be guaranteed, just as in strictly sequential models. But unlike sequential models, competition models like Shortlist do not have problems with an input like *a Christmas pudding*. In Shortlist, activation of *Christmas* will occur at the /k/, and will contribute, via inhibitory competition, to recognition of *a* and rejection of, for example, *acrylic*. This simple example illustrates the most important feature of competition models: Even though direct inhibition occurs only between overlapping candidates, recognition of one word (e.g. *Christmas*) can influence recognition of other words (e.g. *a*) that do not overlap in the input. Inhibitory competition thus provides a means by which words beginning at the same and at different points in the input can be evaluated relative to each other. This relative evaluation process gives competition models a way of segmenting continuous speech.

Although Marslen-Wilson (1987) argued that there was much in common between the Cohort model and the TRACE model, recent accounts (Marslen-Wilson, 1993; Marslen-Wilson & Warren, 1994) have highlighted a fundamental difference between the two models. Marslen-Wilson (1993, p. 205) proposes that competition does not involve any form of lateral inhibition: "Competition ... has no consequences for the relative levels of activation of the competing items ... Activation level ... is determined *only* by degree of bottom-up match or mismatch" (emphasis added). In the revised Cohort model, then, competition is not seen as an active process through which candidate words directly influence the activation levels of each other. Evidence for different candidate words is evaluated separately, with the activation levels of different candidates being determined by the extent to which they match or mismatch with the acoustic–phonetic information in the input. Recognition then depends on a decision stage, where the differential activation levels of the candidates are compared. Recognition of one word requires the differentiation of that word's activation level from that of its competitors. If two candidate words are very similar, the system will take longer to distinguish between them, not because these words inhibit each other directly, but because the decision mechanism will have to wait longer for disambiguating information to arrive which will push the candidates' activation levels apart (Marslen-Wilson, 1993, pp. 205–206). Competition in this model is thus only indirect, or passive: competitors do not compete with each other; their relative merits are instead compared at a decision stage.

As we have argued, direct inhibitory competition offers a way of segmenting continuous speech. This is because it provides a means by which aligned and misaligned lexical hypotheses can be compared. Although the problems of continuous speech recognition have not been addressed in recent descriptions of the Cohort model, it appears that since inhibitory competition has been ruled out, the only mechanism by which aligned and misaligned lexical hypotheses could be compared would be the decision process.

Before the relative strengths of different mechanisms for the evaluation of aligned and misaligned lexical hypotheses should be discussed, it is necessary to establish whether in fact there is a need for any kind of lexical competition. In this paper, we describe two ways to analyse the need for this mechanism. First, its usefulness can be evaluated by examining how vocabulary is organised. In particular, we examine the extent of lexical embedding; that is, how frequently words are embedded in other words. Second, experimental investigation can reveal whether or not lexical competitors influence spoken word recognition. In the second part of the paper, we review the evidence on competition between both aligned and misaligned candidate words.

WORDS WITHIN WORDS

In one part of our Joint Councils Initiative project, we used a large machine-readable dictionary to explore the pattern of occurrence of words within other words in the English vocabulary. This kind of computational analysis provides a means of measuring the extent to which listeners are likely to be faced with particular problems in speech input. If certain patterns of input (such as embeddings) occur frequently in the input, then models that cope well with such input are obviously to be preferred over those which do not.

Lexical competition provides a means of dealing with words embedded in other words. It deals with words embedded at the beginning, in the middle, or at the end of longer words. For example, the embedded words *cat*, *a* and *log* (among others) would enter into the lexical inhibition process in Shortlist, along with the intended word, given the input *catalogue*. But competition ensures that *catalogue*, since it is consistent with a larger portion of the input string, will finally have the highest activation level, and is therefore the most likely candidate to be recognised. A lexical competition process is more plausible if many embedded words are found in the vocabulary, especially if the words involved are themselves frequent in the language. If lexical candidates consistent with the speech input are activated during recognition, words within other words would be highly problematic for recognition, unless a mechanism exists to deal with them. The usefulness of lexical competition between aligned competitors would thus increase if there are many words embedded at the onsets of longer words, and the usefulness of competition between misaligned candidates would grow with the number of words embedded in the middles or the ends of longer words.

We report here the results of several analyses of lexical embedding, as measured in a large machine-readable dictionary of English. The analyses used a lexical database developed from the *Longman Dictionary of Contemporary English* (LDOCE; Alshawi, Boguraev, & Carter, 1989; Carroll, 1992; McQueen & Briscoe, 1991; Procter, 1978). Word-within-word analyses were performed in two stages. First, the phonological strings for each headword between two and six syllables in length were extracted and listed as "matrix" words. Multi-word or phrasal headwords (e.g. *funny peculiar*) were not included in the matrix-word lists. Second, the matrix words were searched for words within them. The searches were based on syllabic matches: a word was considered to be an embedded word only if it matched perfectly the syllabification of the matrix word, as determined by a syllable parser (Carter, 1989) which utilised the phonotactic constraints described by Gimson (1980), plus the Maximal Onset Principle (Selkirk, 1984). Thus syllabic matches (*can* in *canvas*) were counted, while non-syllabic matches (*can* in *scandal*) were ignored. The syllabic constraint rules out some spurious types of embedding that would match phonemically, such as *can* in *scandal* (where the aspiration required for initial /k/ in *can* would be absent in the /sk/ cluster), and *elf* in *shellfish* (where the embedding spans a syllable boundary).

All individual syllables and all sub-strings of syllables within the matrix words were analysed, covering all possible locations of embedded words. Several classes of headwords were excluded: prefixes, suffixes, letters of the alphabet, combining forms (e.g. *-latry*) and apostrophised forms (e.g. *'re*). The embedded words found were paired with their matrix word and listed in output files. Several analyses were then performed on these data. A report of embedded word analyses based on LDOCE, distinguishing between words beginning with strong syllables and words beginning with weak syllables, appears in McQueen and Cutler (1992).

Frequency of Words within Words

The first analysis counted the number of words in the output files which contained no embedded words, the number which contained only one embedded word, and the number containing multiple embedded words. Altogether, 83.8% of English polysyllables were found to contain at least one embedded word, and 63.2% of the polysyllables contained more than one embedded word. Word-within-word embedding is rife in the English vocabulary, such that erroneous lexical candidates will be accessed in over four-fifths of the polysyllabic vocabulary.

Position of Words within Words

Table 1 shows the distribution of embedded words broken down by location and by length of matrix word. In this analysis, we counted a word as having an embedded word in a given location irrespective of the number of words found at that location (i.e. finding *sew*, *so*, *soh* and *sow* in *sodium* counted only once, as did finding *can* in *canvas*); the results thus indicate a lower limit for the frequency of occurrence of words within words. This conservative count nevertheless reveals that there are many words forming the first syllable of longer words: 57.5% of all polysyllables have at least one word embedded as their initial syllable. There are also many words which form the first and second syllables of longer words. The fact that embedding is even more frequent towards the beginning of polysyllabic words than towards their end can be seen from Table 1; even in the longest words, it is the first syllable which is most likely to exist as an independent monosyllabic word. In three- and four-syllable words, it is also the case that embedded two-syllable words are likely to span the first and second syllables of the matrix word. As was described in McQueen and Cutler (1992), the vast majority of these two-syllable embedded words begin with strong syllables.

TABLE 1
Proportions of Polysyllabic Words in the *Longman Dictionary of Contemporary English* (Matrix Words) with at Least One Embedded Word in a Given Location, for Each Matrix Word Length[a]

Matrix Word Length	*Number of Syllables in Embedded Word*	*Location of Onset of Embedded Word in Matrix Word*					
		First Syllable	*Second Syllable*	*Third Syllable*	*Fourth Syllable*	*Fifth Syllable*	*Sixth Syllable*
Two syllables	1	0.651	0.515				
Three syllables	1	0.525	0.467	0.386			
	2	0.246	0.158				
Four syllables	1	0.465	0.388	0.446	0.255		
	2	0.140	0.105	0.142			
	3	0.091	0.098				
Five syllables	1	0.480	0.394	0.393	0.480	0.151	
	2	0.144	0.074	0.302	0.109		
	3	0.058	0.035	0.110			
	4	0.060	0.103				
Six syllables	1	0.535	0.407	0.483	0.459	0.424	0.128
	2	0.186	0.081	0.244	0.070	0.070	
	3	0.087	0.029	0.029	0.052		
	4	0.058	0.017	0.192			
	5	0.017	0.093				

[a]The data are given separately for each embedded word length, and by the location of the onset of the embedded word. For example, *canvas* contributes to the 65% of two-syllable matrix words with a monosyllabic embedded word (*can*) beginning from its first syllable.

Benefits of the Content/Function Distinction

Lexical embedding is a severe problem, particularly at word onsets. It remains possible, however, that syntactic or semantic constraints could rule out a large number of lexical embeddings. For example, content and function words could be distinguished by the recognition system, and thus function words embedded in content words would not pose a problem. It has been proposed (even as far back as Thorne, Bratley, & Dewar, 1968) that distinguishing content from function words could facilitate parsing. Cutler and Carter (1987) have shown that the vast majority of content and function words can be distinguished on phonological grounds (most of the content words in a corpus of conversational English were found to begin with strong syllables, whereas most of the function words could be realised as weak syllables). A content/function word distinction could therefore be a useful heuristic for lexical access algorithms (Cutler & Carter, 1987) and human speech recognition (Cutler, 1993). Separating the functional portion of the vocabulary from the main lexical stock (content words such as nouns, verbs and adjectives) entails, for instance, that recognition of the noun *sodium* will involve no interference from the embedded function word *so* (but note that there would still be interference from the content words *sew*, *soh* and *sow*).

The next two analyses examined whether distinguishing between content and function words would reduce the lexical embedding problem. These analyses replicated the preceding two, but only nouns, verbs, adjectives and adverbs were included in the polysyllabic base-lists. Furthermore, only embedded words which were themselves content words were counted. Auxiliary verbs and the copula *be*, as function words, were also excluded.

Exclusion of function words makes little difference to the overall proportions of words within words. Altogether, 77.9% of polysyllabic content words were found to have at least one embedded content word; 50.3% of all content words were found to have more than one embedded content word. Thus a recogniser using a lexicon of only content words would still generate many erroneous hypotheses.

Table 2 shows the distribution of embedded content words by location. As in the earlier positional analysis (Table 1), we counted the occurrence of at least one word in a given location. This measure of the lower limit of the extent of lexical embedding shows that the exclusion of function words again makes little difference to the statistics. Altogether, 49.7% of polysyllabic content words have at least one content word embedded as their first syllable; and again, embedding is more frequent towards word beginnings than towards word ends.

Benefits of Grammatical Category Constraints

It has been claimed (e.g. Marslen-Wilson & Tyler, 1980) that already recognised syntactic context can be exploited in a top-down fashion to constrain the acceptability of alternative word candidates (though these particular authors have subsequently rejected this type of top-down processing; see, e.g. Marslen-Wilson, 1987). With top-down constraints, a contextually inappropriate embedded word would not interfere with recognition of its matrix word. Where the syntactic context requires an adjective, for instance, *blowzy* should be recognised without interference from the incompatible embedded noun *blouse*. In our final analysis, we compared the grammatical categories of matrix and embedded words to investigate whether incorporating top-down syntactic constraints would indeed nullify the effect of embedded words within words.

We exhaustively compared all of the possible grammatical categories of a matrix word with all of the possible categories of all of the embedded words found at the onset of that word. Included in the category of words embedded in onsets were monosyllables forming the initial syllable of other words, disyllables found in the first and second syllables of words of three or more

TABLE 2
Proportions of Polysyllabic Content Words in the *Longman Dictionary of Contemporary English* (Matrix Words) with at Least One Embedded Content Word in a Given Location, for Each Matrix Word Length[a]

Matrix Word Length	*Number of Syllables in Embedded Word*	*Location of Onset of Embedded Word in Matrix Word*					
		First Syllable	*Second Syllable*	*Third Syllable*	*Fourth Syllable*	*Fifth Syllable*	*Sixth Syllable*
Two syllables	1	0.594	0.408				
Three syllables	1	0.434	0.310	0.321			
	2	0.238	0.156				
Four syllables	1	0.364	0.236	0.315	0.199		
	2	0.130	0.101	0.137			
	3	0.091	0.098				
Five syllables	1	0.380	0.250	0.260	0.349	0.131	
	2	0.123	0.068	0.282	0.105		
	3	0.057	0.034	0.111			
	4	0.059	0.102				
Six syllables	1	0.453	0.291	0.372	0.256	0.331	0.105
	2	0.145	0.070	0.244	0.064	0.070	
	3	0.087	0.029	0.029	0.052		
	4	0.058	0.017	0.192			
	5	0.017	0.093				

[a]The data are given separately for each embedded word length, and by the location of the onset of the embedded word. For example, *canvas* contributes to the 60% of two-syllable matrix words with a monosyllabic embedded word (*can*) beginning from its first syllable.

syllables, and so on, up to any five-syllable words which were the first five syllables of six-syllable words.

This analysis showed that grammatical category constraints failed to remove the interference from embedded words, although they did succeed in reducing it: two-thirds of all words embedded in the onset of other words (66.2%) had a grammatical category differing from that of the matrix word. Therefore, although grammatical category constraints reduce the problem of words within words, by no means do they remove it. A substantial number (i.e. 33.8%) of all embedded words do match the grammatical category of their matrix word (i.e. cannot be excluded by a grammatical filter). Interestingly, the proportion is similar for subcategories within embedded words; thus, of the monosyllables which form the initial syllable of other words, 33.6% match the matrix word category, and hence cannot be excluded by a grammatical filter.

LEXICAL EMBEDDING AND RECOGNITION MODELS

The above analyses make competition models of spoken word recognition highly plausible. Overall, one can expect that four out of five English polysyllables will contain at least one embedded word. Lexical embedding is so common in the vocabulary that most words will have many potential competitors.

The simplest grammatical constraint, distinguishing content from function words, fails to reduce the number of words within words. Stringent grammatical category filtering does improve the situation, but even under such constraints one-third of embedded words at the beginnings of longer words cannot be rejected. It is worth noting that grammatical filters can only operate with maximum efficiency if the preceding context uniquely determines the grammatical class of the word being recognised. This is rarely the case. For example, although the context may determine that the next word must be the beginning of a noun phrase, determiners, adjectives, pronouns, proper names, mass nouns and plural count nouns could all appear in this position. The recognition system certainly could not depend on the preceding context to provide sufficient disambiguating information (Norris, 1982). And even if such a filter were to operate perfectly, embedded words of the same category as the longer word would still be accessed.

It remains possible that the 33% of embedded words which match the matrix word syntactically could perhaps be excluded on the basis of semantic constraints. This cannot be checked automatically using the LDOCE lexical database, but in any case, the same limitations apply to a semantic filter as apply to a syntactic filter. Preceding context does not reliably determine the semantics of the following word. This is not to say that syntactic and semantic information is not used during word recognition. It is well-established that recognition is both faster and more accurate when contextual information is available (Bard et al., 1988; Marslen-Wilson & Tyler, 1980). In the Shortlist model (Norris, 1994a), context can boost the activation levels of plausible candidates and decrease the activation of implausible candidates; however, as in other models (e.g. the Cohort model; Marslen-Wilson, 1987), contextual information is not available prelexically, so cannot directly determine either which candidate words are accessed, or at which points in the input access attempts should be made. There are therefore constraints on the role context plays in recognition and segmentation; in recent models, it only operates after lexical access has taken place, and, in any case, it cannot always provide sufficient disambiguating information to uniquely determine appropriate lexical hypotheses.

All of our analyses used a very tight phonological constraint. Only embedded words matching the syllabification of the matrix word were counted. Furthermore, only one syllabification of both the embedded word and the matrix word was considered (as defined by Gimson, 1980; Selkirk, 1984). These analyses thus underestimate the proportion of lexical embeddings in the vocabulary. Higher numbers of lexical embeddings will obviously be found when the embedded word only needs to match the matrix word on segments, irrespective of syllabification. Frauenfelder (1991), who computed embedding on the basis of segmental matches, indeed found a high degree of lexical embedding in the Dutch vocabulary. The overall patterns remained the same, however—words were again most frequently found at the onsets of longer words. Analyses performed on a corpus of *spoken* English (MARSEC; Roach, Knowles, Varadi, & Arnfield, 1993) confirm that the pattern of embedding is very similar for segmental- and syllabic-based matches (Cutler, McQueen, Baayen, & Drexler, 1994). These corpus-based analyses naturally integrate effects of relative word frequency, and are thus similar to frequency-weighted analyses of the vocabulary. They therefore also show that the patterns reported here are not simply due to embedding in rare words which listeners do not normally hear. In other words, the proportion of lexical embedding found in the vocabulary is reflected in actual spoken language.

The data of Cutler et al. (1994) also indicate that although the positional patterns remain the same, there are many more embedded words found when matching is based on segmental information alone than when it is also based on syllabic information. As Briscoe (1989) has argued, using syllabic information will constrain the number of words that the recognition system needs to consider. But as our current analyses indicate, the syllabic constraint fails to remove the problem of words within words.

Our results highlight the need for a lexical competition process. They also further emphasise the inadequacy of strictly sequential models, underlining why they have been abandoned by psycholinguistic theory. The efficiency of such models depends both on the number of words embedded at the onsets of longer words, and on the number of words embedded medially and finally in longer words. The high incidence of word-onset embedding is a problem for sequential models because successful recognition entails correct rejection of onset embeddings. For example, the recognition of *operative* will be affected by the onset embedding *opera*. It is possible that these models would falsely recognise such embedded words. It is certainly the case that recognition of *operative* could only be achieved after the candidate *opera* had been rejected. The large numbers of word-internal and word-final embeddings found are also problematic. If sequential recognition were to fail, word onsets would be postulated at incorrect positions. The onset of an embedded word could be postulated wrongly, and that embedded word could then be recognised. For example, if the first two syllables of *candidate* were recognised incorrectly as *candy*, a word boundary would be postulated after the /I/, and then *date* would also be recognised incorrectly. Strictly sequential models therefore appear ill-suited to the structure of the English vocabulary.

The vocabulary is such that a mechanism by which aligned lexical hypotheses can be compared seems essential. Word onsets so frequently have words embedded within them that successful recognition must depend on an efficient means of dealing with these embeddings. Similarly, embedded words appear regularly in non-onset (misaligned) positions, so the recognition system should also have a means of rejecting these words. This argument for lexical competition between both aligned and misaligned lexical hypotheses would be strengthened if we could show that listeners are influenced by competition: Do competitor words actually affect recognition performance?

RECOGNITION BY COMPETITION

There is a growing body of experimental evidence for competition in spoken word recognition. Studies using the cross-modal semantic priming task (Swinney, 1979) have provided evidence that multiple lexical hypotheses are activated (Marslen-Wilson, 1987; 1990; Shillcock, 1990; Swinney, 1981; Zwitserlood, 1989). If subjects hear part of a word that is consistent with both that word and another lexical hypothesis, they are faster in making visual lexical decisions to semantic relatives of both the actual word and the alternative candidate (Zwitserlood, 1989). Thus, subjects were faster to decide that both *schip* (ship) and *geld* (money) were words (in Dutch), on hearing up to the /t/ in *kapitein* (captain). It appears that both *kapitein* and *kapitaal* (capital) were activated in this situation. Likewise, lexical decisions were faster to words like *rib*, presented at the offset of *trombone*, suggesting that *bone* was activated (Shillcock, 1990). These results are consistent with competition models, but provide no direct support for a competition process, since they indicate only that multiple hypotheses are activated, not that they compete with one another.

More direct evidence for competition between activated word candidates has been obtained in tasks examining phonological priming in monosyllabic words (Goldinger, Luce, Pisoni, & Marcario, 1992; Slowiaczek & Hamburger, 1992). Listeners find it harder to recognise a target word (as measured by lexical decision, repetition, or identification in noise) when it has been preceded by a prime word sharing phonetic or phonological material. These inhibitory effects in priming (as opposed to some facilitatory effects which can be attributed to strategic processes) have been taken to be due to competition between activated lexical hypotheses. The prime activates words in the same phonological space (i.e. words overlapping in form, including the target), which then compete for recognition. Subsequent recognition of the target is then more difficult, either because the entire word set is more active, or because the target word has residual inhibition. Both of these accounts of the priming effect require that similar words are in competition during word recognition. Other results from the priming paradigm show effects of the number of words activated by a given input (neighbourhood effects), and effects of the relative frequencies of occurrence of the target and the set of competitors (Goldinger, Luce, & Pisoni, 1989; Luce, Pisoni, & Goldinger, 1990).

Cluff and Luce (1990) have found evidence of competition in bisyllabic items consisting of two monosyllabic words (such as *madcap*). The monosyllables were either easy to recognise (they were high-frequency words in sparse, low-frequency neighbourhoods, such as *mad*) or hard to recognise (they were of low frequency in dense, high-frequency neighbourhoods, such as *cap*). Listeners were asked to detect the bisyllables in white noise. Several competition effects were found. For example, words with a hard first syllable were harder to identify when the second syllable was also hard than when it was easy. Words embedded in other words, and the phonetic neighbours of those words, appear to be activated and to compete for recognition with the embedding word.

Although the above research suggests that multiple words are simultaneously active, and that there is competition between these word hypotheses, it does not distinguish between models like Shortlist, in which competition occurs through lateral inhibitory connections between candidate words, and other models—like the Cohort model—in which competition operates at a decision stage (Marslen-Wilson, 1993). Much of the above research has in fact been interpreted as support for the Neighborhood Activation Model (NAM; Luce et al., 1990), in which competition is also decision-based. In the NAM, recognition is determined by a relative-goodness decision rule, in which the evidence in favour of a word is compared with the evidence in favour of other words activated by the same stretch of input. Competitors do not inhibit each other directly, as in Shortlist, but there is a process of relative evaluation of competitors at the decision stage. These results do not distinguish between inhibitory and decision-based competition because they refer to aligned competition, and to the recognition of isolated words. The two alternatives can only be distinguished when misaligned competitors, and the problems of continuous speech recognition, are considered.

SEGMENTATION BY COMPETITION

An important part of our JCI project was to test experimentally the claims about lexical competition made in the Shortlist model. The research outlined above only provides evidence for competition between competitors which are aligned in the input. We therefore investigated whether competition operates not only between aligned competitors but also between competitors which are misaligned.

McQueen et al. (1994) asked subjects to detect real words (such as *mess*) embedded in nonsense strings which were either the onsets of longer real words (such as /dəmɛs/, the onset of *domestic*) or not (such as /nəmɛs/, which cannot be continued to form any words). The subjects found it harder to detect the target words in the word onsets. The Shortlist model predicts this competition effect between misaligned candidates. The evidence for the target word begins to arrive later than that for the competitor (for example, *domestic* begins to receive activation from the onset of /dəmɛs/, whereas *mess* is only activated after the arrival of the /m/). The model also predicts an effect when the target and competitor are fully aligned (for example, *sack* and *sacrifice* given /sækrəf/), as was found (for example, word-spotting was easier in /sækrək/ than /sækrəf/). Finally, the model predicts that the competition effect should be stronger when the target starts later than the competitor than when the words are aligned (because in the former case the competitor can be more inhibiting, since it has an opportunity to build up activation before the target becomes activated). The competition effects were indeed larger in the items like /dəmɛs/ than in the items like /sækrəf/.

Norris et al. (in press) extended these results by showing that there are effects of inhibitory competition where evidence for the competitors starts later than that for the target. Cutler and Norris (1988) showed that subjects found it more difficult to spot, for example, *mint* at the onset of a string consisting of two strong syllables (SS, e.g. /mɪnteɪv/) than at the onset of a string with a strong followed by a weak syllable (SW, e.g. /mɪntəv/). It was claimed that the reason for this is that the second strong syllable in an SS string triggers a segmentation process, making a word onset at the beginning of the second syllable (the /t/ in /mɪnteɪv/) more likely. This interferes with recognition of the target (*mint*), and slows its detection, relative to the SW case where no segmentation process is triggered by the second weak syllable. This result is consistent with others in the literature (Cutler & Butterfield, 1992; McQueen et al., 1994), suggesting that in a stress-timed language like English, strong syllables have a special role to play in speech segmentation. What Norris et al. (in press) showed was that the interfering effect of the second syllable in SS strings is modulated by the number of competitors beginning at that syllable. The difficulty of word spotting in SS relative to SW strings was greater for targets like *mask* in strings like SS /m*a*skʌk/ and SW /m*a*skək/, where there were many words beginning at the second syllable (from the /k/), than for targets like *mint* in strings like SS /mɪntaʊp/ and SW /mɪntəp/, where there are few second syllable competitors (beginning from the /t/).

It is worth noting that these results indicate that segmentation of continuous speech is achieved by the combined effects of lexical competition and metrical structure. Cross-linguistic research has shown that different units are involved in the segmentation of different languages (the strong syllable in English: Cutler & Norris, 1988; the syllable in French: Cutler, Mehler, Norris, & Segui, 1986; 1992; Mehler, Dommergues, Frauenfelder, & Segui, 1981; and the mora in Japanese: Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993). These units each contribute to the metrical structure of the relevant language. It thus appears that across languages, metrical information is used in speech segmentation. But our results (McQueen et al., 1994; Norris et al., in press) have shown that metrical segmentation is not sufficient for continuous spoken word recognition—lexical competition also plays a role. We have also shown that these two mechanisms are entirely compatible—metrical segmentation for English can be implemented in the Shortlist model (Norris et al., in press). Strong syllables are given a special role to play in the computation of lexical activation levels (activation is boosted in words which are aligned with strong syllables, and activation is penalised in words which are misaligned with strong syllables), and then, as before, these candidate words compete for recognition. Accurate simulation of our results did not fall out of the competition process alone, even with a large lexicon where any asymmetries between strong and weak syllables would be represented. But the data could be modelled successfully when metrical information in the input was allowed to influence the activation of lexical hypotheses. The activation/competition mechanism in Shortlist thus provides a means for the instantiation of metrically based segmentation processes.

In Dutch, Vroomen and de Gelder (1995) have also shown effects of the number of words beginning from the second syllable of bisyllabic strings, using a cross-modal identity priming task. Listeners were faster in visual lexical decision to *melk* (milk) after hearing *melkem* (no second syllable competitors beginning from the /k/) than after hearing *melkeum* (few second syllable competitors). These responses to *melk* were in turn faster than those made after hearing *melkaam* (many second syllable competitors), which in turn were faster than those made after hearing a control word *lastem*. Lexical candidates activated by input arriving later in time than the target word influenced recognition of that target. Vroomen and de Gelder (1995) interpreted their results in terms of the Shortlist model.

Recent research has therefore shown that competition operates both between lexical hypotheses that begin at the same point in the speech signal and between hypotheses which begin at different points. Competition is therefore a mechanism which serves two functions in continuous speech understanding: the recognition of words and the segmentation of the input.

The inhibitory competition mechanism instantiated in the Shortlist model is able to perform these two functions, as our simulations show.

Marslen-Wilson (1993) has argued, on the basis of a number of priming experiments, that competition does not take the form of lateral inhibition between activated hypotheses. Recognition of, for example, PATCH was primed by prior presentation of an associate such as *cabbage*, but PATCH was not reliably primed by a closely mismatching word such as *cabin*. Similarly, a word like WOUND was primed by *bandage* but not by a mismatching nonword such as *bandin*. The failure to obtain priming by closely mismatching items suggests that activation levels of candidate words are highly sensitive to the degree of acoustic–phonetic match between the input and lexical hypotheses. The mismatching final /n/ in *cabin*, for example, appears to be enough to penalise severely the activation level of *cabbage*. Marslen-Wilson (1993) argues further that lateral inhibition predicts that mismatch effects should be more severe when the mismatch creates another word than when it creates a nonword (since in addition to a poorer bottom-up match of, for example, *cabin* to *cabbage*—equivalent to the *bandin–bandage* case—the activated candidate *cabin* should also penalise through lateral inhibition the activation level of *cabbage*). The failure to obtain differential mismatch effects might therefore be seen as evidence against competition models.

This argument, however, depends on the relative effects on activation levels of bottom-up mismatch information and of lexical competition. If bottom-up mismatch alone is sufficient to devastate the activation of a candidate word, then it may be impossible to see any further mismatch effects due to competition. If the activation levels of, for example, *cabbage* and *bandage* were both effectively set to zero by a mismatching final /n/, then it would be impossible for any additional lateral inhibitory effect of *cabin* to differentiate the near-zero activation levels of *cabbage* and *bandage*. The Shortlist model, in contrast to TRACE, employs the use of bottom-up mismatch information to penalise mismatching candidate words very strongly. A mismatching phoneme penalises the activation level of a candidate word three times more than a matching phoneme boosts a candidate's activation level (Norris, 1994a). Shortlist can thus easily accommodate the results of Marslen-Wilson (1993).

GENERAL DISCUSSION

The lexical statistics we have presented indicate that polysyllabic words in the English vocabulary usually have shorter words embedded within them, and that a lexical competition mechanism would have considerable scope to operate. In addition to the recent literature suggesting that aligned lexical hypotheses compete with one another (as in the case of isolated word recognition), we have presented results showing that there is also competition between aligned and misaligned candidates. The experimental results and the lexical statistics are both problematic for strictly sequential models. We now return to the issue we first raised in the Introduction, namely whether competition should be instantiated by a direct inhibitory competition process, or by a decision mechanism.

Some authors (Bard, 1990; Marslen-Wilson, 1993) have asked whether it is possible to distinguish between inhibitory competition and competition due to decision processes. Decision mechanisms are certainly able to account for competition effects. For example, the presence of two overlapping candidates could slow recognition simply because a decision mechanism delayed responding until one of the candidates could be eliminated. Such a decision mechanism could operate by monitoring activity in the competing candidates and responding as soon as evidence for one candidate was sufficiently greater than evidence for other candidates. With such a system, competitors would slow recognition even without any direct inhibition between competing candidates.

In fact, both the decision and the inhibitory forms of competition can be incorporated into the same model (Norris, 1994b). However, although competition effects can emerge from a decision mechanism rather than from inhibition, competition by inhibition is itself a form of decision mechanism. This is particularly apparent in the existing computational implementations of TRACE and Shortlist on serial computers. A more fruitful contrast between different forms of competition can be achieved by considering the size and location of units that enter into the competition process.

For competition to provide a means of segmenting continuous speech, as well as recognising individual words, words beginning at different points in the signal must be able to be compared. One means of achieving this would be if alternative analyses of stretches of the input were compared. A decision-based competition model would have to operate in this way, by comparing the accumulating evidence for alternative strings of candidate words. But these large units would be inefficient. If alternative analysis paths of a long input were compared, the number that the recognition mechanism would have to consider could potentially become enormous, and the same words, in different parses, might have to be represented several times. Harrington and Johnstone (1987) have shown that phonemic transcriptions of some short utterances (less than 10 words) can be parsed into more than 10,000 different strings of words. This problem could be solved utilising a network-based structure in which distinct portions of such paths were represented only once. This, of course, is exactly the solution used by competitive inhibition models—each individual word is represented once, no matter how many possible analyses it may be involved in.

Without inhibitory competition, however, there is no way a system comparing words rather than analysis paths can show appropriate sensitivity to non-overlapping input. Consider how a decision-based system would respond to the input *ship inquiry*. The requirement to select the best overlapping competitor at each point would force the system to choose *shipping* rather than *ship*. To appreciate that *shipping* should be considered less likely because it also overlaps with *inquiry*, the system would need to take account of exactly the same information as inhibitory competition models like Shortlist and TRACE do. The system would need to be able to compare misaligned words (*shipping* and *inquiry*) and, most importantly, to use the fact that *shipping* overlaps with *inquiry* to make *shipping* a less effective competitor. The inhibitory competition mechanism in Shortlist ensures that the activation of each candidate is sensitive to the impact which that candidate has on the interpretation of both that part and other parts of the utterance. A decision-based mechanism would also have to show this kind of sensitivity. Both mechanisms have to be able to weigh up each candidate with respect not just to that candidate's fit to the part of the input with which it is aligned, but also with respect to how that candidate fits with other candidates, spanning other parts of the input. A mechanism basing decisions on individual words would thus be functionally indistinguishable from an inhibitory competition mechanism.

Relative evaluation of aligned and misaligned hypotheses is the essence of inhibitory competition models. The evidence in favour of each word must be a function of the bottom-up support for that word and the degree to which it overlaps with other candidates. The effect of overlapping candidates must itself be modulated by the strength of overall evidence for those candidates. This is readily conceptualised in terms of inhibitory connections between word nodes as in Shortlist and TRACE. However, as Norris (1994a) points out, other forms of constraint satisfaction mechanisms could equally well perform the same function without the need to employ explicit inhibitory connections. In any mechanism, whether or not it actually incorporates inhibitory *connections*, overlap between any pair of competitors has to inhibit or reduce their activation or probability of recognition in such a way that both of those competitors will have less strength to compete with other words. The central claim of an inhibitory competition model like Shortlist is a psychological claim about the function and dynamics of the computations being performed, not about the manner in which those computations are implemented.

While inhibitory competition could be thought of as being a function of a decision process, however, not all decision-based competition systems will show the essential properties of inhibitory competition systems. Simpler decision models without competitive inhibition might possibly be adequate for visual word recognition where word boundaries are known in advance and the task is simply to determine which single word in the lexicon best fits a given letter string, but they are incapable of operating in continuous speech, which has no reliably marked word boundaries. Only competition by inhibition can perform the dual functions of recognition and segmentation required for spoken word recognition. This is because competition allows both aligned and misaligned words to be compared.

In conclusion, recognition and segmentation of continuous speech appear to be based on competition between lexical hypotheses. Competition models like Shortlist (Norris, 1994a) are able to deal with the large proportion of embedded words in the English vocabulary. Furthermore, our experimental evidence (McQueen et al., 1994; Norris et al., in press) supports the direct lexical competition instantiated in the Shortlist model. We have shown that the recognition of a word is inhibited by the previous activation of competitors preceding the target (where recognition of *mess* in /dəmɛs/ is made more difficult by the activation of *domestic*), by the concurrent activation of a competitor aligned with the target (where recognition of *sack* in /sækrəf/ is made more difficult by the activation of *sacrifice*), and by the subsequent activation of competitors following the target (where recognition of *mask* in /m*a*skʌk/ is influenced by the number of words beginning with the /k/). It appears that both aligned and misaligned candidates compete, and that competition is therefore involved in both recognition and segmentation. Lexical competition seems to be a necessary mechanism for the recognition of words in continuous speech.

REFERENCES

Alshawi, H., Boguraev, B., & Carter, D.M. (1989). Placing the dictionary on-line. In B. Boguraev & E.J. Briscoe (Eds), *Computational lexicography for natural language processing*, pp. 41–64. London: Longman.

Bard, E.G. (1990). Competition, lateral inhibition, and frequency: Comments on the chapters of Frauenfelder and Peeters, Marslen-Wilson, and others. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, pp. 185–210. Cambridge, MA: MIT Press.

Bard, E.G., Shillcock, R.C., & Altmann, G.T.M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics*, *44*, 395–408.

Briscoe, E.J. (1989). Lexical access in connected speech recognition. In *Proceedings of the Association of Computational Linguistics*, Vancouver, June, pp. 84–90.

Carroll, J. (1992). The ACQUILEX lexical database system: System description and user manual. In A. Sanfilippo (Ed.), *The (other) Cambridge Acquilex papers*. Technical Report TR-253, Computer Laboratory, University of Cambridge.

Carter, D. (1989). LDOCE and speech recognition. In B. Boguraev & E.J. Briscoe (Eds), *Computational lexicography for natural language processing*, pp. 135–152. London: Longman.

Cluff, M.S., & Luce, P.A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 551–563.

Cole, R.A., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing*, pp. 67–116. London: Academic Press.

Cole, R.A., & Jakimik, J. (1980). A model of speech perception. In R.A. Cole (Ed.), *Perception and production of fluent speech*, pp. 133–163. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research*, *22*, 109–131.

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 and Language*, *2*, 133–142.

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

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

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., McQueen, J., Baayen, H., & Drexler, H. (1994). Words within words in a real-speech corpus. In *Proceedings of the 5th Australian International Conference on Speech Science and Technology*, Perth, December, Vol. I, pp. 362–367.

Frauenfelder, U.H. (1991). Lexical alignment and activation in spoken word recognition. In J. Sundberg, L. Nord, & R. Carlson (Eds), *Music, language, speech and brain*, pp. 294–303. Wenner-Gren International Symposium Series. London: Macmillan.

Gimson, A. (1980). *An introduction to the pronunciation of English*, 3rd edn. London: Edward Arnold.

Goldinger, S.D., Luce, P.A., & Pisoni, D.B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*, 501–518.

Goldinger, S.D., Luce, P.A., Pisoni, D.B., & Marcario, J.K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 1211–1238.

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

Harrington, J., & Johnstone, A. (1987). The effects of equivalence classes on parsing phonemes into words in continuous speech recognition. *Computer Speech and Language*, *2*, 273–288.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, *51*, 2018–2024.

Luce, P.A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics*, *39*, 155–158.

Luce, P.A., Pisoni, D.B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, pp. 122–147. Cambridge, MA: MIT Press.

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

Marslen-Wilson, W.D. (1990). Activation, competition, and frequency in lexical access. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, pp. 148–172. Cambridge, MA: MIT Press.

Marslen-Wilson, W.D. (1993). Issues of process and representation in lexical access. In G.T.M. Altmann & R. Shillcock (Eds), *Cognitive models of speech processing: The second Sperlonga meeting*, pp. 187–210. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.

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

Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.

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

McQueen, J.M., & Briscoe, E.J. (1991). A computational tool for examining lexical segmentation in continuous speech. In *Proceedings of EUROSPEECH '91*, Geneva, September, Vol. 2, pp. 697–700.

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

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.H., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*, 298–305.

Nakatani, L.H., & Dukes, K.D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, *62*, 714–719.

Norris, D.G. (1982). Autonomous processes in comprehension: A reply to Marslen-Wilson and Tyler. *Cognition*, *11*, 97–101.

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

Norris, D.G. (1994b). A quantitative multiple-levels model of reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1212–1232.

Norris, D., McQueen, J.M., & Cutler, A. (in press). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258–278.

Procter, P. (Ed.) (1978). *Longman dictionary of contemporary English*. London: Longman.

Roach, P., Knowles, G., Varadi, T., & Arnfield, S. (1993). MARSEC: A machine-readable spoken English corpus. *Journal of the International Phonetics Association*, *23*, 47–53.

Selkirk, E.O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Shillcock, R.C. (1990). Lexical hypotheses in continuous speech. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, pp. 24–49. Cambridge, MA: MIT Press.

Slowiaczek, L.M., & Hamburger, M.B. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 1239–1250.

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

Swinney, D.A. (1981). Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In T. Myers, J. Laver, & J. Anderson (Eds), *The cognitive representation of speech*, pp. 201–209. Amsterdam: North-Holland.

Thorne, J., Bratley, P., & Dewar, H. (1968). The syntactic analysis of English by machine. In *Machine Intelligence 3*. New York: Elsevier.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 98–108.

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