

Leading Up the Lexical Garden Path: Segmentation and Ambiguity in Spoken Word Recognition

Matthew H. Davis and William D. Marslen-Wilson
Medical Research Council Cognition and Brain Sciences Unit

M. Gareth Gaskell
University of York

Two gating studies, a forced-choice identification study and 2 series of cross-modal repetition priming experiments, traced the time course of recognition of words with onset embeddings (*captain*) and short words in contexts that match (*cap tucked*) or mismatch (*cap looking*) with longer words. Results suggest that acoustic differences in embedded syllables assist the perceptual system in discriminating short words from the start of longer words. The ambiguity created by embedded words is therefore not as severe as predicted by models of spoken word recognition based on phonemic representations. These additional acoustic cues combine with post-offset information in identifying onset-embedded words in connected speech.

An important problem in the perception of connected speech is segmentation: how listeners divide the speech stream into individual lexical units or words. Words in fluent speech are not separated by silence in the same way that printed words are divided by blank spaces, yet connected speech is perceived as a sequence of individual words. This perceptual experience clearly reflects acquired language-specific knowledge, because listeners do not have the same experience when hearing an unfamiliar language.

Theories of word segmentation distinguish between processes that operate on either a prelexical or a lexical representation of the speech signal (Gow & Gordon, 1995). At the prelexical level, a variety of cues have been proposed, allowing segmentation through the use of acoustic cues to word onsets (Lehiste, 1960; Nakatani & Dukes, 1977) or from knowledge of statistical regu-

larities of lexical items (such as distributional regularity, phonotactics, or metrical stress; see Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Cutler & Norris, 1988). However, because not all words can be segmented in this way, accounts of spoken word recognition also incorporate mechanisms by which lexical identification can contribute to speech segmentation. Two main lexical accounts of segmentation have been proposed—the sequential recognition of words in connected speech (Cole & Jakimik, 1980; Marslen-Wilson & Welsh, 1978) or lexical competition between word candidates spanning word boundaries (McClelland & Elman, 1986; Norris, 1994).

In this article we will consider a topic that has long been considered to favor lexical competition accounts: the presence of words that are embedded at the onset of longer words (such as *cap* embedded in *captain* or *captive*). Many models of segmentation and lexical access predict ambiguity between short and long words during the processing of longer words that contain embeddings. This temporary ambiguity may challenge sequential recognition accounts of segmentation (Luce, 1986) and has been argued to support accounts of segmentation based on interword competition (McQueen, Cutler, Briscoe, & Norris, 1995).

The goal of our research was to take a more detailed look at the recognition of short words (*cap*) that are embedded in longer competitors (*captain*). We used gating and cross-modal priming to measure the time course by which these items were activated and identified in sequences of connected speech. First, we compared sentences containing short words with matched sequences containing longer words in which the short words were embedded. This allowed us to measure the ongoing ambiguity between short and long words for sequences in which either word was viable. Reliable differences in the articulation of syllables in short and long words have long been noted (Klatt, 1976; Lehiste, 1972). In our experiments, we tested whether these acoustic differences play a role in distinguishing onset-embedded words from longer competitors. A further set of investigations tested a specific prediction of lexical competition accounts—that the identification of short words would be disrupted by sequences in which the following

Matthew H. Davis and William D. Marslen-Wilson, Medical Research Council (MRC) Cognition and Brain Sciences Unit, Cambridge, United Kingdom; M. Gareth Gaskell, Department of Psychology, University of York, Heslington, United Kingdom.

This research was supported by a United Kingdom Engineering and Physical Sciences Research Council research studentship awarded to Matthew H. Davis and by a United Kingdom Medical Research Council grant awarded to William D. Marslen-Wilson and Lorraine K. Tyler. Results of Experiments 1 and 2 were presented at the 19th annual conference of the Cognitive Science Society, July 1997, Palo Alto, California, and at the January 1997 meeting of the Experimental Psychology Society, London, United Kingdom.

We are grateful to Ingrid Johnsrude, Kathy Rastle, Lolly Tyler, Billi Randall, Tom Loucas, and members of the Centre for Speech and Language, Department of Experimental Psychology, Cambridge University, Cambridge, United Kingdom, for comments and suggestions on this work. We are also grateful to James McQueen and two anonymous reviewers for comments on previous drafts of this article. Thanks to Brian Cox for assistance with the figures.

Correspondence concerning this article should be addressed to Matthew H. Davis, MRC Cognition and Brain Sciences Unit, Chaucer Road, Cambridge CB2 2EF, United Kingdom. E-mail: matt.davis@mrc-cbu.cam.ac.uk

context for these words creates a *lexical garden path*, because the subsequent word temporarily matches a longer word (e.g., in the sequence *cap tucked*, the /t/ in *tucked* is also found in the longer word *captain*). By comparing garden-path and non-garden-path sequences, we evaluated the role of post-offset information in the identification of short words that are embedded in longer competitors.

Although our research is mainly concerned with lexical accounts of segmentation, we subsequently review prelexical accounts, because the absence of robust prelexical cues has motivated theories in which lexical knowledge is used in the segmentation of connected speech.

Prelexical Accounts of Segmentation

Acoustic Cues to Word Boundaries

Investigations of minimal pairs that differ only in the location of a word boundary (such as *play taught* and *plate ought*) have identified a variety of acoustic cues that are associated with word boundaries (Lehiste, 1960). In English, allophonic differences in the articulation of segments at the onset of words (such as glottal stops, laryngeal voicing, and aspiration) can be used by listeners as cues to word boundaries (Christie, 1974; Nakatani & Dukes, 1977). Word-initial segments are also longer in duration than equivalent segments that are not word initial (Lehiste, 1960). However, measurements of connected speech have shown that these cues may be less clearly marked or nonobligatory in more naturally produced stimuli (Barry, 1981).

Another prelexical cue to segmentation is duration differences between segments and syllables in short and long words (Klatt, 1976; Lehiste, 1972; Umeda, 1975). For instance, the syllable /sliɪp/ is progressively shortened in the words *sleepy* and *sleepiness* compared with the monosyllable *sleep* (Lehiste, 1972). Listeners have been shown to use these changes in duration in segmenting sequences of repeated syllables (Nakatani & Schaffer, 1978). However, because other factors such as phonetic, prosodic, discourse level, and interspeaker variation also affect the duration of segments and syllables, duration cues may not be sufficiently reliable to allow the segmentation of connected speech (for further discussion of sources of variation in segment and syllable duration, see Anderson & Port, 1994; Crystal & House, 1990; Klatt, 1976).

Distributional Accounts of Segmentation

Distributional theories of segmentation have proposed that the speech stream can be divided into words, prior to lexical access, by using knowledge of the statistical properties of lexical items. Analyses of phonemically transcribed corpora have shown a variety of regularities that may be used to predict the location of word boundaries. For instance, the metrical segmentation strategy (MSS) proposed by Cutler and colleagues (Cutler & Butterfield, 1992; Cutler & Norris, 1988) among others (Grosjean & Gee, 1987) suggests that because the majority of English content words start with a strong syllable (Cutler & Carter, 1987), stressed syllables should be treated as potential word onsets. In accordance with this idea, experiments using word spotting have shown that listeners are faster to detect monosyllables followed by strong syllables than those followed by weak (unstressed) syllables (Cut-

ler & Norris, 1988; Norris, McQueen, & Cutler, 1995; Vroomen, van Zon, & de Gelder, 1996).

Another distributional account assumes that chunking the speech stream into frequently occurring phoneme sequences will produce segmentation into linguistically coherent units (Brent & Cartwright, 1996; Perruchet & Vinter, 1998; Wolff, 1977; see Harris, 1955, for a related approach in linguistics). Experiments in which adults are familiarized with nonword sequences support this account (Dahan & Brent, 1999). The reverse approach—that infrequent sequences are assumed to straddle word boundaries—has also been proposed. A variety of symbolic and connectionist systems have demonstrated that this form of phonotactic knowledge provides a plausible prelexical segmentation strategy (Aslin, Woodward, LaMendola, & Bever, 1996; Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Christiansen, Allen, & Seidenberg, 1998; Gaskell, 1994; Harrington, Watson, & Cooper, 1989). In support of these accounts, word-spotting experiments have shown that words bounded by phonotactically illegal sequences are detected more easily than words ending in a legal sequence (in adults, McQueen, 1998; in 9-month-old infants, Mattys, Jusczyk, Luce, & Morgan, 1999). Saffran and colleagues (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) demonstrated that both adults and 8-month-old infants use transitional probabilities to segment artificial speech after only a few minutes' exposure. A further distributional cue is the *possible word constraint*, that all words segmented out of connected speech must include a vowel (Brent & Cartwright, 1996; Norris, McQueen, Cutler, & Butterfield, 1997). This constraint improves the segmentation performance of a competition-based recognition model, matching the results of word-spotting and priming experiments.

However, none of these distributional accounts are sufficiently reliable to allow all segmentation to occur prelexically. Even when multiple sources of information from phonotactics and metrical stress are combined (as in the neural network simulations reported by Christiansen et al., 1998), distributional information locates less than half of all word boundaries. Indeed, the more successful models that incorporate distributional cues use a lexicon of frequently occurring sequences to assist segmentation (see Brent, 1999, for a review). The acoustic cues to word boundaries that we described previously may assist the operation of these distributional systems. However, authors have typically dismissed these acoustic cues as too unreliable to be useful (Brent & Cartwright, 1996) or have treated natural variation in the speech stream as a potentially problematic source of noise (Christiansen & Allen, 1997). Consequently, the distributional accounts described thus far typically operate on phonemically labeled input. The limited success of these accounts, combined with the apparent unreliability of acoustic cues, perhaps explains the prevalence of accounts of segmentation that use lexical information.

Lexical Accounts of Segmentation

Lexical accounts of word segmentation propose that the identification of individual lexical items contributes to the placement of boundaries between words. These accounts can be divided into two main classes: those that propose segmentation is achieved by the sequential recognition of individual words in the speech stream (e.g., Cole & Jakimik, 1980; Marslen-Wilson & Welsh, 1978) and

those that propose segmentation arises through competition between lexical items that cross potential word boundaries (e.g., McClelland & Elman, 1986; Norris, 1994).

Sequential Recognition

An influential lexical account of segmentation was proposed as part of the Cohort model of spoken word recognition (Marslen-Wilson & Welsh, 1978). One property of the Cohort theory is that the recognition system processes incoming speech in a maximally efficient manner. Multiple lexical candidates are activated by the initial portion of a word and are progressively eliminated by mismatching input until a single lexical item remains. Thus, words are identified as soon as the speech stream uniquely specifies a single lexical item. This proposal has received empirical support from results showing that words with an early uniqueness point (i.e., words for which the sequence of sounds diverges from other lexical items early in the word) are identified more rapidly than are words with a late uniqueness point (Marslen-Wilson, 1984, 1990; Zwitserlood, 1989).

If words can be recognized before their acoustic offset, then lexical knowledge can be used to predict the location of upcoming word boundaries. However, investigations of large lexical databases show that many words do not diverge from all other candidates until after their final segment (Luce, 1986; McQueen et al., 1995). These are words that are embedded at the onset of longer lexical items (such as *cap* embedded in *captain* or *captive*). Embedded words are problematic for sequential recognition accounts such as the Cohort model because these accounts do not allow for cases in which it may not be possible to identify words before their acoustic offset.

Recurrent network simulations (Gaskell & Marslen-Wilson, 1997b; Norris, 1990) that implement a sequential recognition process make this limitation of the original cohort account particularly transparent.¹ At the end of a sequence of phonemes (such as /kæp/), the output of the network is in an ambiguous state with both the short word (*cap*) and longer word candidates activated (*captain*, *captive*, etc.). For sequences in which /kæp/ comes from a short word (such as *cap* in the sequence *cap fits*) the network will activate a new set of candidates beginning with the segment /f/ (*feel*, *fall*, *fit*, etc.) at the onset of the following word. Therefore, at no point does the network resolve the ambiguity between embedded words and the start of longer competitors.

The prevalence of onset-embedded words (Luce, 1986; McQueen et al., 1995) is therefore argued to rule out any account of spoken word recognition that is reliant on preoffset identification. Consequently, models of spoken word recognition propose that lexical processing must continue across word boundaries and that competition between lexical candidates that span word boundaries is used to segment connected speech.

Lexical Competition

Competition between activated candidates is commonly invoked as a mechanism for lexical selection. It is proposed that if multiple lexical items are activated, then the activation of each item will be reduced. Models of spoken word recognition that incorporate lexical competition include the neighborhood activation model (Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990), the

revised Cohort model (Marslen-Wilson, 1987, 1990; Marslen-Wilson, Moss, & van Halen, 1996), and connectionist models such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994). Although these models make different proposals regarding the nature of the competitor environment and whether competition effects arise at the lexical level or in subsequent decision stages, all agree that lexical activation will be reduced when a greater number of lexical candidates match the speech input.

Effects of lexical competition have been shown in a number of studies. For instance, the number and frequency of phonetic neighbors has been shown to influence the speed of responses in lexical decision and shadowing, as well as influencing the percentage of correct identifications in noise (Luce & Pisoni, 1998; Luce et al., 1990). Evidence of lexical competition has also come from cross-modal priming studies that show reduced facilitation when prime stimuli remain ambiguous between multiple lexical candidates (Gaskell & Marslen-Wilson, 1997a; Marslen-Wilson et al., 1996; Moss, McCormick, & Tyler, 1997; Zwitserlood & Schriefers, 1995). Finally, in auditory lexical decisions, previous presentations of words, such as *bruise*, slow subsequent responses to onset-aligned competitors, such as *broom* (Monsell & Hirsh, 1998). The increased activation of a phonologically similar item can delay recognition of a competitor.

These experiments have demonstrated competition between related items during single-word processing (when word boundaries are explicitly specified). However, in some accounts, overlapping items in connected speech are also lexical competitors, despite not sharing the same word boundaries (McClelland & Elman, 1986; Norris, 1994). Competition between overlapping candidates results in a lexical system that will settle into a state representing a consistent segmentation of a sequence. Thus, lexical competition provides a mechanism for lexical segmentation as well as for lexical selection.

Evidence of competition between nonaligned lexical hypotheses comes from word-spotting and cross-modal priming experiments reported by Vroomen and de Gelder (1995; see also Norris et al., 1995). Vroomen and de Gelder found that activation of Dutch words such as *melk* (milk) are reduced in sequences such as *melkaam* compared with sequences such as *melkeum*. Crucial to this is the fact that the second syllable *kaam* forms the start of many Dutch words, whereas *keum* is a less frequent word onset. Thus, competition from multiple candidates starting with *kaam* reduces the activation of the overlapping word *melk*.

Two models of spoken word recognition that incorporate lexical competition between nonaligned candidates are TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997). For instance, these models would include an inhibitory connection between the lexical unit for *cap* and all other words that contain the sequence /kæp/. These inhibitory connections will allow the identification of onset-embedded words. For example, in the sequence *cap fits*, information after the end of the word *cap* can rule out longer competitors,

¹ These recurrent network simulations extend the original cohort account, because the activation of lexical candidates represents the conditional probability of each word given the current input. In the original Cohort model described by Marslen-Wilson and Welsh (1978), activation of word candidates is binary.

such as *captain*, and boost the activation of embedded words through reduced competitive inhibition. Thus, lexical competition can assist the identification of short embedded words by allowing information after the offset of a word to rule out longer competitors.

Simulations using TRACE have demonstrated that post-offset mismatch and lexical competition allows the identification of onset-embedded words (Frauenfelder & Peeters, 1990). Words embedded in longer competitors are identified in cases in which mismatch occurs immediately after the offset of the embedded word (as in *cap fits*) or for sequences in which mismatch is delayed until after word offset (e.g., in lexical garden-path sequences such as *cap tucked*). Similar effects of competition and delayed mismatch have been described in Shortlist (McQueen, Norris, & Cutler, 1994).

However, the delayed identification that is predicted by lexical competition accounts may not be the only means by which embedded words can be identified. If additional information is present in the speech input to distinguish short words from the start of longer competitors, delayed identification may be unnecessary. Because the presence of embedded words has motivated competition-based accounts of lexical segmentation, we subsequently review the experimental literature for evidence of the delayed recognition predicted by these accounts.

Recognizing Embedded Words

Experiments using the gating task, such as Grosjean (1985), have shown that low-frequency monosyllables in noninformative sentences may not be isolated or recognized until after their acoustic offset. For stimuli taken from conversational speech, Bard, Shillcock, and Altmann (1988) reported that 20% of words (mostly short, closed-class words) are not identified before their acoustic offset. Although some incorrect responses in Grosjean's study were longer words in which the target was embedded (e.g., the response *bunny* for the target word *bun*), neither of these studies specifically chose words that were embedded in longer words.

The recognition of embedded words has been studied more directly using word spotting (Cutler & Norris, 1988; McQueen, 1996; McQueen et al., 1994). McQueen et al. (1994) demonstrated that the detection of monosyllables in bisyllabic strings is more difficult if the stimulus is a fragment of a longer word. Detection responses were slower for the word *mess* in the sequence /dəmes/ (from *domestic*) than in the nonword sequence /nəmes/. Similar results were obtained for onset-embedded words; participants found it harder to detect *sack* in /sækrəf/ (from *sacrifice*) compared with the nonword sequence /sækrək/, though this effect was only apparent through increased error rates. These findings provide evidence for effects of competition between embedded words and longer lexical candidates.

Experiments reported by Luce and Lyons (1999) investigated the recognition of longer words that contain embeddings, using lexical decision and shadowing. Latencies for words that contained an onset-embedding were faster than those for matched words without an embedded word. This facilitation runs contrary to the predictions of competition-based models. However, because the methods used by McQueen et al. (1994) and Luce and Lyons provide only an indirect measure of the activation of short and long

word candidates, competition effects on lexical activation cannot be distinguished from effects at decision stages that may be differentially involved in word spotting or lexical decision.

One way to assess the activation of words in connected speech is through the cross-modal priming of lexical decision responses (Swinney, Onifer, Prather, & Hirshkowitz, 1979). Experiments using cross-modal priming have demonstrated that alternative segmentations of connected speech are activated during recognition. For instance, Shillcock (1990) demonstrated significant priming of the target word *RIB* by sentences containing the word *trombone* (through the embedded word *bone*). This has been confirmed using single word presentations (Isel & Bacri, 1999; Luce & Cluff, 1998; Vroomen & de Gelder, 1997; however, see Marslen-Wilson, Tyler, Waksler, & Older, 1994). These results suggest that nonaligned lexical hypotheses are activated during connected speech. However, comparisons of the activation of appropriate and inappropriate segmentations of offset-embedded words (Isel & Bacri, 1999) showed greater priming of targets related to the longer word (*trombone*) than of targets related to the embedded word (*bone*), suggesting that the perceptual system is able to distinguish between offset-embedded words and longer competitors by the end of a word.

Isel and Bacri (1999) also showed an absence of priming for words related to onset embeddings (e.g., *BUS* related to the word *car* in *cargo*). However, because these experiments only probed at the offset of the bisyllable, it is unclear whether transient activation of onset-embedded words might have been observed at earlier probe positions. Experiments by Tabossi, Burani, and Scott (1995) demonstrated equal priming for associates of the word *visite* (visit) from sentences containing *visite* and from sequences in which *visite* was formed by sections of two adjacent words (as in the sequence *visi tediati* [faces bored]), but did not investigate the activation of short words that are embedded in longer items (e.g., probing *visi* during *visite*).

Another study that compared activations of both appropriate and inappropriate segmentations of embedded word stimuli was carried out by Gow and Gordon (1995). Results showed priming of the target *KISS* from related two-word stimuli (*two lips*), though not from sentences containing the phonemically identical (but acoustically distinct) single word (*tulips*). Conversely, targets related to the long word (*FLOWER*) were primed by both single-word (*tulips*) and two-word (*two lips*) stimuli. Gow and Gordon concluded that these results indicate sensitivity to acoustic cues that mark word onsets.

In summary, several studies have demonstrated ambiguity created by the lack of marked word boundaries in connected speech. However, only two studies (Gow & Gordon, 1995; Isel & Bacri, 1999) have assessed the severity of this ambiguity by testing the ability of listeners to distinguish appropriate from inappropriate segmentations of words containing embeddings. Such comparisons are necessary to determine whether the acoustic cues previously described play a role in the perception of embedded words in connected speech. One limitation of these studies is that they do not provide information on the time course of identification of embedded words, preventing direct comparison with models of spoken word recognition that simulate this time course. Experiments that are able to trace the identification of onset-embedded words and longer competitors across measured portions of speech would therefore inform our understanding of the cues and mech-

anisms that are used for the identification of onset-embedded words in connected speech.

Outline of This Research

The primary goal of our research was to pin down the precise degree of online ambiguity between short words and the onsets of longer words in which these short words are embedded. The greater the ambiguity, the more plausible are proposals for delayed recognition through lexical competition. We tested this proposed ambiguity by tracking the activation of onset-embedded words and longer competitors, using sentence fragments cut off at positions of interest in the speech stream. By comparing the activation profile of short words and longer competitors, we were able to evaluate how different sources of information in the speech stream contribute to the identification of these items.

We used two main methods to evaluate the activation of competing interpretations of our test sequences. In Experiments 1 and 4, we used a gating task in which we presented participants with progressively longer fragments of speech and they wrote down the words that they could identify at each gate. In Experiments 2 and 5, we used cross-modal repetition priming to provide an online measure of lexical activation at each probe position. In Experiment 3, we used a forced-choice gating experiment to follow up on a crucial finding from Experiments 1 and 2.

Our initial series of experiments (Experiments 1, 2, and 3) maximized potential ambiguity and delayed recognition through the presentation of short words in *lexical garden-path sequences*. These are sequences in which the continuation of an embedded word temporarily matches a longer competitor (such as the segment /t/ in *cap tucked*, which matches the competitor *captain*). For these sequences, coarticulatory influences from the following syllable would not rule out the presence of a longer word. Furthermore, the presence of segments that match a longer word would increase the activation of longer competitors, further depressing the activation of the embedded word. The activation of short and long words for these lexical garden-path sequences was compared with matched sentences containing a longer lexical item in which the short word was embedded (e.g., *captain*).

In a second series of experiments, we investigated the more usual case in which continuations of the short word immediately mismatch with longer competitors (such as the sequence *cap looking*, in which the initial syllable plus continuation [kæpl] is inconsistent with longer lexical items). Comparison between these two sets of sequences allowed us to test the effect of information coming after the offset of embedded words—in particular whether competition from longer items suppresses the activation of onset-embedded words. Below, we describe the word pairs and sentence contexts that were used throughout these experiments.

Materials

Forty bisyllabic words were selected from the CELEX lexical database (University of Nijmegen, Nijmegen, Holland; see Baayen, Piepenbrock, & Guilikers, 1995) that had an unrelated monosyllable phonologically embedded at their onset (e.g., *captain* containing the word *cap*). All bisyllables had a full first syllable and an embedded monosyllable with at least three letters and three phonological segments matching the syllabification of

the longer word. Items were rejected if they could not be used in the same syntactic class,² or if they were orthographically unusual (e.g., *pizza*). However, not all items were orthographically as well as phonologically embedded (pairs such as *track* and *tractor* were included). Monosyllables were embedded in a limited number of longer words ($M = 12$; maximum = 43; minimum = 1). Short and long words had approximately equal CELEX frequencies (mean frequency: short words = 35/million, long words = 25/million), $t(39) = 1.07$, $p > .1$.

Pairs of monosyllables and bisyllables were placed in nonbiasing test sentences with an average of six syllables of preceding context (Range = 3 to 11 syllables). A cloze test was carried out to ensure that neither word could be predicted from the preceding context. Continuations for the short word stimuli were chosen that had the same onset as the second syllable of the longer word. An example pair of sentences is shown below (with target words underlined):

Short word—*The soldier saluted the flag with his cap tucked under his arm.*

Long word—*The soldier saluted the flag with his captain looking on.*

The 40 pairs of sentences (shown in the Appendix) were recorded onto digital audiotape in a soundproof booth by Matthew H. Davis. They were recorded successively to ensure that intonation patterns were identical. Care was taken not to include a prosodic boundary after the embedded word, because this would provide a nonlexical cue to the presence of a word boundary. These recordings were passed through an antialiasing filter and digitized at a sampling rate of 22 kHz using a Data Translation DT2821 sound card attached to a Dell PC. The start and the end of each sentence was marked using the Brown Lab Interactive Speech System (BLISS) speech editing system (Mertus, 1990; for more information, see Mertus, 1989). An additional marker was also placed at the onset of the target word—a point at which each pair of sentences should be nearly identical.

Forced-Choice Cloze Pretest

A forced-choice, auditory cloze pretest was carried out to confirm that (a) there was no semantic or pragmatic bias toward short or long words in these context sentences and (b) the onset of the sentence did not contain any cue (in intonation or prosody) to the identity or length of the target word. Forty participants were tested on one version of each of the paired sentences presented, up to the start of the target word. They were instructed to indicate which of the two target words (e.g., *CAP* and *CAPTAIN*) was more likely to occur next. Responses were made on a 6-point scale, indicating overall preference for short or long words and confidence levels.

Results were analyzed both as confidence ratings and as the proportion of responses favoring short and long word continua-

² For one pair (*gray-gravy*), the embedded word was used as an adjective in the sentence. One bisyllable (*can-canteen*) contained two full syllables (with primary stress on the second syllable); all other items had primary stress on their initial syllable. One reviewer suggested that possible differences might exist in embedded syllables for two items (*gin-ginger*, *junk-junction*). Removal of these items did not affect the results obtained in our experiments.

tions (determined by the side of the scale on which responses fell). Across the two sets of 40 sentences, one-sample t tests showed that confidence ratings did not significantly differ from the middle rating of 3.5 ($M = 3.63$; 1 represents a short word and 6 a long word), $t(39) = 0.83$, $p > .1$, and that proportions of long word responses did not differ from 0.5 ($M = 0.543$), $t(39) = 0.99$, $p > .1$. These measures showed that the sentence contexts were not consistently biased semantically or pragmatically toward either test word. Comparisons of responses made to short and long word sequences did not show any significant differences in either confidence ratings (short words = 3.65; long words = 3.61), $t(39) = 0.667$, $p > .1$, or response proportions (short words = 0.549, long words = 0.537), $t(39) = 0.574$, $p > .1$. We concluded that prior to the onset of the target word, the experimental recordings do not contain any acoustic cue that could bias listeners toward short or long words.

Alignment Points and Acoustic Analysis

To ascertain whether acoustic differences existed between embedded syllables in short and long word stimuli and to allow comparisons between stimuli containing equivalent acoustic-phonetic information, we set up *alignment points* at matched positions in each pair of sentences. Acoustic differences and differences in participants' interpretations were measured with respect to three alignment points, which are illustrated for an example stimulus pair in Figure 1.

The first alignment point (AP_1) was placed at the offset of the syllable forming the embedded word, such as /kæp/ in *cap* or *captain* (see Warren & Marslen-Wilson, 1987, for the procedure used to identify syllable offsets). Measurements of the duration of this embedded syllable (onset to AP_1) showed a marked difference in acoustic duration between syllables from short and long words (syllable duration: monosyllables = 291 ms, bisyllables = 243 ms), $t(39) = 9.35$, $p < .001$. The voiced portion of this syllable was extracted using the voicing and pitch detection algorithm supplied in ESPS/Waves (Talkin, 1995). This voiced portion accounted for the majority of the duration difference in these syllables (duration of voicing: monosyllables = 216 ms, bisyllables = 172 ms), $t(39) = 8.50$, $p < .001$. There was also a marginally significant difference in the mean F_0 for this syllable (monosyllables = 112 Hz; bisyllables = 115 Hz), $t(39) = 1.75$, $p < .1$, and no difference in the average amplitude of these syllables (root mean square [RMS] energy: monosyllables = 2,731, bisyllables = 2,755), $t(39) = 0.30$, $p > .1$.

The second alignment point (AP_2) was placed after the onset segments of the second syllable but before steady vowel periodicity was apparent. A paired t test showed that there were no significant differences in the duration of syllable onsets that were word initial in the short word stimuli and word medial in the long word stimuli (onset duration: short word = 72 ms, long word = 77 ms), $t(39) = 0.42$, $p > .1$. This null finding contrasts with the stimuli used by Gow and Gordon (1995) and may be attributable

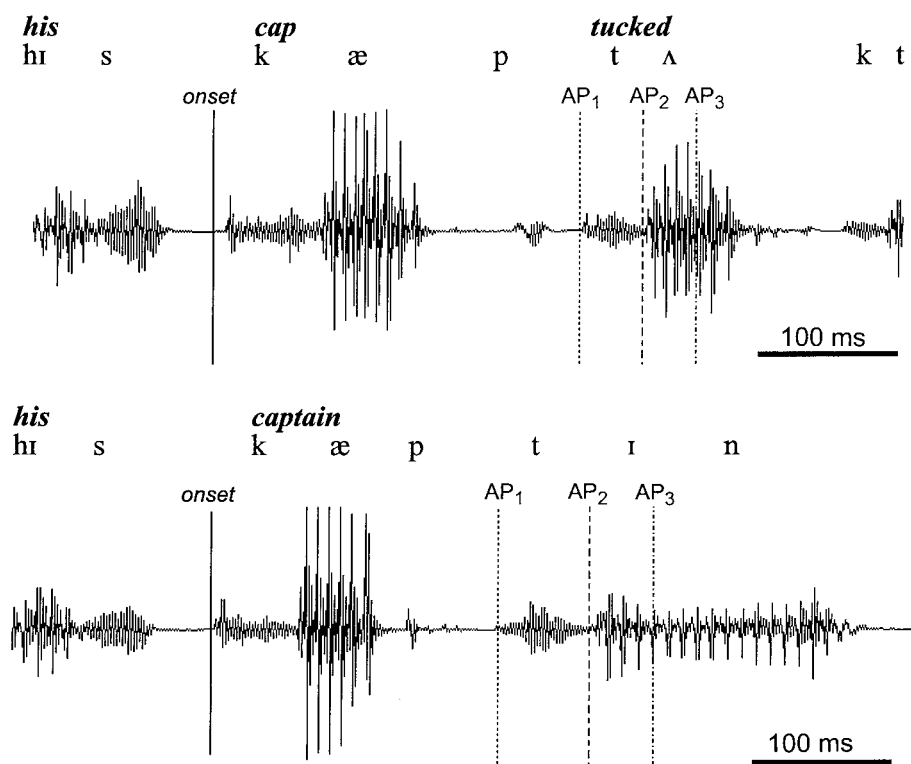


Figure 1. Speech waves and alignment points (APs) for the stimuli in Experiment 1. Marked points correspond to the onset of the target word, the offset of the target word (AP_1), the onset of the second syllable (AP_2), and the vowel of the second syllable (AP_3). Stimulus items are *The soldier saluted the flag with his cap tucked under his arm* (top) and *The soldier saluted the flag with his captain looking on* (bottom).

to our attempts to exclude prosodic boundaries between short words and their continuations. Other research has shown that increases in segment duration at word onset are more marked at prosodic boundaries (Klatt, 1976) and may be less marked in connected speech than in citation forms (Barry, 1981). Nonetheless, the absence of an expected acoustic difference in onset segments may increase the ambiguity of short and long word sequences.

The third alignment point (AP₃) marked the earliest location where the stimuli differ phonemically. This marker was placed four pitch periods into the vowel of the second syllable (duration AP₂ to AP₃: monosyllables = 42 ms, bisyllables = 44 ms), $t(39) = 1.35, p > .1$.

Experiment 1: Gating

We used the gating task (Grosjean, 1980, 1996) to assess the sections of the speech stream that support the identification of embedded words and longer competitors. The critical comparison was whether responses to short and long word stimuli diverge before or after the offset of the embedded syllable. If listeners are sensitive to the acoustic differences in embedded syllables that were described previously, then responses to short and long word stimuli would diverge prior to AP₁. However, if syllables of onset-embedded words create ambiguity, then responses to short and long word stimuli would remain identical until at least the first alignment point (AP₁). Such ambiguity would delay recognition until boundary cues in the onset of the following word (AP₂) or phonemic mismatch in the vowel (AP₃) can be detected.

Method

Participants. Twenty-four participants from the Birkbeck Centre for Speech and Language (CSL) subject pool were tested. Most were University of London students; all were aged between 18 and 45 years and were paid for their participation. All were native speakers of British English with normal hearing and no history of language impairment.

Design and materials. Participants made written responses to successively presented fragments of the test sentences. The dependent variable was the word identified at each gate, and the independent variables were whether the sentence fragments contained a short or a long word and which of the 10 fragments of the sentence was presented. Cut-off points for the sentence fragments were the three alignment points shown in Figure 1, as well as two initial gates, 50 and 100 ms before AP₁, and 5 gates placed 50, 100, 200, 300, and 400 ms after AP₃. It was expected that the word after the target could be identified at the final gate. For each fragment, the sentence was presented from the beginning to the cut-off point. The 40 sentence pairs were pseudorandomly divided into two experimental versions, each containing one of each stimulus pair. An additional 20 sentences were added to each version: four practice items to acquaint participants with the task and 16 fillers to distract from the embedded words in the test sentences.

Procedure. Participants were tested in groups of 2 to 4, in booths in a quiet room. They were provided with answer books containing the sentence onset and were instructed to identify all the words that continued each sentence. Binaural, monophonic sentence fragments were played from a 486 PC equipped with a DT2821 sound card through closed-ear headphones. Fragments were presented successively, with a 6-s interval separating each fragment and an extra 2-s delay at gates after AP₃. The test and filler sentences were divided into four 20-min blocks with breaks provided between blocks.

Results and Discussion

Data from 2 participants were rejected because they did not respond to every fragment presented. The remaining responses were coded for whether the initial word matched either target word. Participants produced correct responses for the majority of the test items by the final gate. Three items (*ban*, *bran*, and *win*) were not recognized by over 50% of participants and were excluded (with the bisyllables *bandage*, *brandy*, and *winter*) from further analysis. The proportions of responses matching the short or long target words are shown in Figure 2.

As is apparent in Figure 2, there was an overall bias toward short word responses at the first three gates (until the offset of the embedded syllable at AP₁). Differences between responses to short and long word stimuli were also apparent at these early gates. ANOVAs were carried out on the proportions of responses (averaged over participants and items) that matched each target word using the repeated measures factors of stimulus type (short or long word) and gate number (Gate 1, 2, or AP₁). Analyses by participants included a repeated measures factor of version, and analyses averaged over items included a factor indicating the item group in the rotation (Pollatsek & Well, 1995). These ANOVAs showed that significantly more short word responses were made to short word stimuli than to long word stimuli, $F_1(1, 20) = 60.32, p < .001$, and $F_2(1, 35) = 26.86, p < .001$. There was also a significant effect of gate, $F_1(2, 40) = 30.37, p < .001$, and $F_2(2, 70) = 9.60, p < .001$, and an interaction between stimulus type and gate significant by participants and not items, $F_1(2, 40) = 5.81, p < .01$, and $F_2(2, 70) = 2.35, p > .1$. A similar pattern was observed for long word responses over the first three gates; participants made significantly more long word responses to long word stimuli than to short word stimuli, $F_1(1, 20) = 7.34, p < .05$, and $F_2(1, 35) = 4.69, p < .05$, with a significant effect of gate, $F_1(2, 40) = 14.40, p < .001$, and $F_2(2, 70) = 8.39, p < .001$, and a marginally significant interaction between stimulus type and gate, $F_1(2, 40) = 6.48, p < .01$, and $F_2(2, 70) = 2.82, p < .1$.

Effects of stimulus type in these analyses suggest that participants used acoustic differences between short and long word sequences in responding to the initial syllables of the target word. Despite the effect of these acoustic differences, however, the recognition of short words was delayed in comparison with the identification of longer words. Only at Gate 8 was an equal proportion of correct responses made to short and long word stimuli, $t(36) = 0.96, p > .1$. This delayed recognition may result from competition from long words, because at AP₂ (the onset of the second syllable), many participants gave long word responses to short word stimuli. Only when there was clear phonemic mismatch between short word stimuli and long target words at AP₃ and beyond were participants able to revise these hypotheses and identify the short words. This result confirms the role of information coming after the offset of a word in identifying embedded words (Bard et al., 1988; Grosjean, 1985).

However, these effects of following context may be exaggerated by a bias toward producing the shortest single word that encompasses all speech that has been heard (Tyler, 1984). This response bias could account for the predominance of short word responses at the initial three gates, leading us to underestimate the effectiveness of the acoustic cues distinguishing short and long words. The increase in long word responses at AP₂ could also be explained by

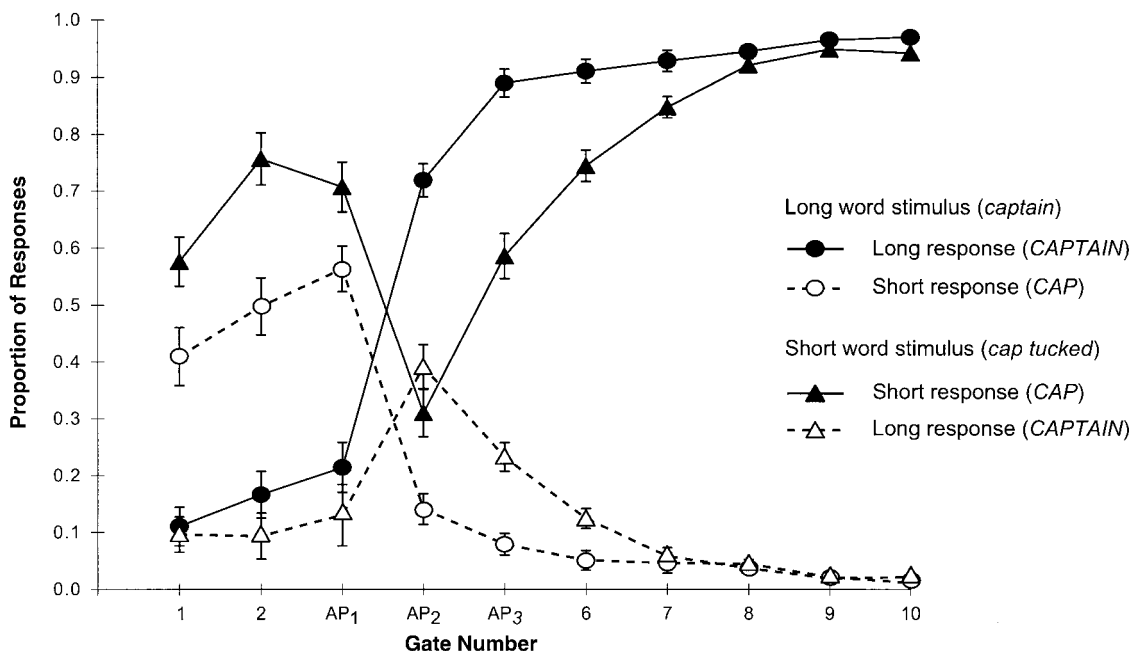


Figure 2. Results for Experiment 1: Proportions of responses at each gate matching short and long target words for stimuli containing short and long words. Error bars show one standard error. AP = alignment point.

a bias toward single word responses. Responding with a single word (*captain*) for stimuli, such as [kæpt] from *cap tucked*, would increase the observed delay in recognizing short words.

Experiment 2: Cross-Modal Priming

The results of Experiment 1 suggest that listeners are sensitive to acoustic differences between syllables of short and long words but that competition from long words might still delay the identification of short word stimuli. Response biases in the gating task, however, may have distorted this picture. It was therefore important to obtain converging evidence in support of these findings. Online methods of probing lexical activation are particularly valuable in this respect, because we can expect them to be less susceptible to biases toward short or long lexical items. Therefore, in Experiment 2, we used cross-modal priming to measure the activation of short and long word interpretations of sequences containing embedded words and longer competitors.

Cross-modal priming of lexical decision responses is a well-established method for probing the activation of potentially ambiguous sequences (Gow & Gordon, 1995; Swinney et al., 1979). Target words are commonly semantically or associatively related to the meaning of the prime word (see, e.g., Luce & Cluff, 1998; Moss & Marslen-Wilson, 1993; Shillcock, 1990; Zwitserlood, 1989; Zwitserlood & Shriefers, 1995; see Tabossi, 1996, for a review). However, in cases in which alternative interpretations are orthographically distinct, it is also possible to use repetition priming with visual targets that are identical to the auditory test prime. Although this method might be expected to be susceptible to form-based facilitation, experiments that include related nonword trials show that repetition priming provides a measure of lexical activation and competition that is comparable with semantic or

associative priming (Gaskell & Marslen-Wilson, 1996, 1997a; Marslen-Wilson, 1990; Marslen-Wilson, Nix, & Gaskell, 1995; Vroomen & de Gelder, 1995; see Zwitserlood, 1996, for a review). Because the competing interpretations in our study have distinct orthographic forms, we could use repetition priming to gauge lexical activation. This provides a robust priming effect and avoids possible confounds produced by differences in semantic or associative relatedness, while retaining sensitivity to effects of lexical competition and mismatch.

In these experiments, spoken primes were cut off at varying points in the test sentences. This allowed us to trace the time course of lexical activation while controlling how much of the prime stimulus was presented (cf. Zwitserlood, 1989). Because the test stimuli and cut-off points were identical to those used in gating, results obtained could be compared with those from Experiment 1. In the initial priming experiment, we used stimuli cut off at the offset of the embedded syllable (AP₁). Gating results indicated that cues exist to distinguish syllables from short and long words, though response biases might have reduced the sensitivity of the gating task to detect influences of these cues.

Experiment 2A

Method

Participants. We tested 74 paid participants from the Birkbeck CSL subject pool, none of whom had taken part in Experiment 1.

Design and materials. The 40 pairs of test sentences from Experiment 1 were used, presented up to the offset of the initial syllable of the target word (AP₁; /kæp/ from *cap* or *captain*). The auditory prime was followed by a visually presented short or long target word for a lexical decision response. Response times (RTs) following test sentences were compared with RTs following unrelated control primes. Control prime sentences were

identical to the test sentences in all but the word at the probe position, which was replaced with a contextually appropriate monosyllable or bisyllable matched in frequency to the short test word (see Table 1). By comparing RTs following the two different control primes, we could determine whether visual lexical decisions following the initial syllable of a bisyllabic control word show significant disruption, compared with targets presented at the offset of a monosyllable.

The experimental design therefore consisted of four prime types (two test and two control primes) paired with either a short or a long target word, producing the eight conditions shown in Table 1. The 320 test trials (40 items in eight conditions) were rotated into eight experimental versions, such that each participant heard only one version of each sentence and responded to only one member of each pair of target words. An additional 80 filler sentences were interspersed with the test items. Of these fillers, 20 sentences were followed by a nonword target that was phonologically similar to the prime word at the probe position. These filler items were added to discourage participants from associating phonological overlap with a *yes* response. The other 60 fillers were followed by unrelated word and nonword targets, such that there was an even number of word and nonword targets in the experiment. Twenty practice items and 10 lead-in items were also added, resulting in experimental versions of 150 items. Of these items, 20 were stimuli in which a phonologically related word target was presented after the auditory prime (13% of all trials). To encourage participants to attend to the prime sentences, we gave a recognition test on some of the filler sentences at the end of the experimental session.

Procedure. Participants were tested on one of the eight experimental versions in groups of 1 to 4. They were warned that they would be given a memory test on the auditory stimuli, but instructed not to rehearse or memorize the sentences. The experiment was split into four sessions: a practice session followed by two blocks of test items, finishing with the pencil-and-paper recognition test.

Stimulus presentation and response measurement was controlled by DMASTR and VMASTR software (Forster & Forster, 1995) on a 486 PC running MS-DOS. Each trial started with a sentence fragment played over headphones. At AP₁ for the test and control prime sentences or at a range of positions in the filler sentences, the speech was cut off and a word presented for 200 ms on a computer monitor. Participants made a speeded lexical decision response using their dominant hand for the *yes* button. RTs were measured from the onset of the target word (the offset of the prime stimulus), with a 3-s time-out. Following the participant's response, there was a short pause before the start of the next trial. Each test session, including practice items and recognition test, lasted approximately 25 min.

Results

Of the 74 participants, the data from 9 were excluded because of slow or error-prone lexical decision responses (mean word RT

greater than 750 ms or lexical decision errors greater than 12.5%). One test item (*bran*) produced a large number of errors (over 30%) and consistently slowed RTs (over 750 ms); it was removed from further analysis, along with its associated bisyllable (*brandy*). Four outlying data points over 1,200 ms were also excluded (0.16% of the data). Mean RTs and error rates are shown in Table 2, with the key priming effects plotted in Figure 3A.

Given our goal of using the magnitude of priming as an index of lexical activation, we focused statistical analyses on whether responses to short and long target words were facilitated by test primes compared with responses following control primes. Before carrying out these analyses, however, we needed to determine the appropriate baseline from which to measure priming effects. Therefore, we compared RTs following the two control prime conditions to determine whether cutting off a prime stimulus partway through a word had any effect on subsequent lexical decisions.

Two-way ANOVAs were carried out on response means following control primes, averaged over participants and items, using the repeated measures factor of target length (short vs. long target words) and prime length (short vs. long control prime stimuli). We included test version as a variable in these analyses to reduce estimates of random variation caused by differences between participants and items assigned to each experimental list (Pollatsek & Well, 1995). In the analysis by participants, this factor referred to the test version to which each participant was assigned. In the item analysis, version refers to the eight groups of items sharing the same (pseudorandom) assignment of conditions to test versions. As discussed by Pollatsek and Well (1995), main effects and interactions involving version and item-group factors are indicative of between-groups variability; therefore, we do not report them.

As is apparent from the condition means shown in Table 2, there was no significant difference between RTs following a short or long control prime ($F_1 < 1$ and $F_2 < 1$). Thus, lexical decision responses were not affected by cutting off primes in the middle of a word compared with cutting off speech at the offset of a word. Responses were significantly faster for short target words compared with long target words, $F_1(1, 57) = 49.90, p < .001$, and $F_2(1, 31) = 25.71, p < .001$, whereas the interaction between prime and target length was nonsignificant ($F_1 < 1$ and $F_2 < 1$). Given the lack of significant differences between short and long

Table 1
Prime and Target Stimuli for Experiment 2

Prime type	Prime stimulus	Short target	Long target
Short test	<i>The soldier saluted the flag with his <u>cap</u>^a <u>t^bu^cck^ded</u> under his arm.</i>	CAP	CAPTAIN
Long test	<i>The soldier saluted the flag with his <u>cap</u>^a<u>t^ba^cn^d</u> looking on.</i>	CAP	CAPTAIN
Short control	<i>The soldier saluted the flag with his <u>palm</u> facing forwards.</i>	CAP	CAPTAIN ^e
Long control	<i>The soldier saluted the flag with his <u>rifle</u> by his side.</i>	CAP ^e	CAPTAIN

Note. AP = alignment point.

^a Approximate probe position for Experiment 2A (AP₁). ^b Approximate probe position for Experiment 2B (AP₂). ^c Approximate probe position for Experiment 2C (AP₃). ^d Approximate probe position for Experiment 2D (AP₃ + 100 ms). ^e Condition was excluded from Experiments 2B, 2C, and 2D.

Table 2
Mean Response Times (RTs; in Milliseconds) and Error Rates (in Percentages) by Prime and Target Type for Experiment 2A

Prime type and word	Short target (CAP)		Long target (CAPTAIN)	
	RT	Error	RT	Error
Short test (<i>cap</i>)	485	3.2	539	4.8
Long test (<i>captain</i>)	501	2.8	528	6.4
Short control (<i>palm</i>)	512	3.8	561	4.3
Long control (<i>rifle</i>)	512	2.8	557	6.0

control primes, we collapsed these two conditions in subsequent analyses.

We examined the significance of priming effects observed in this experiment by carrying out planned pairwise comparisons between RTs following test and (pooled) control primes (see Monsell & Hirsh, 1998, for a similar approach to the analysis of priming experiments). Following Toothaker (1991), we carried out these comparisons using repeated measures ANOVAs, including version and item–group factors as before. To control familywise error rate for these comparisons, we used the modified Bonferroni correction for multiple planned comparisons proposed by Keppel (1982, p. 148). This controls familywise error rate to a level determined by the number of degrees of freedom in the original data set. Our data set comprised two separate families of comparisons for short and long targets. For each family of comparisons, we had three conditions (three prime types) for which we conducted three comparisons. The modified Bonferroni correction, therefore, requires that comparisons pass a threshold of $p < .033$ to be reported as reaching a corrected $p < .05$.³ The magnitude and significance of comparisons between RTs following test and control primes are plotted in Figure 3A.

Strongest priming effects in this experiment were observed when the prime syllable came from the same word as the target. Lexical decision responses to long target words were most strongly facilitated by long word primes, $F_1(1, 57) = 20.09, p < .001$, and $F_2(1, 31) = 24.30, p < .001$, with weaker and marginally significant facilitation following short word primes, $F_1(1, 57) = 6.57, p < .1$, and $F_2(1, 31) = 4.31, p < .1$. Responses to long word targets were also significantly faster following related long word primes than when following related short word primes, $F_1(1, 57) = 4.72, p < .05$, and $F_2(1, 31) = 5.03, p < .05$. Conversely, responses to short target words were significantly faster following related short word primes than when following control primes, $F_1(1, 57) = 15.13, p < .001$, and $F_2(1, 31) = 23.23, p < .001$, whereas short word responses were not significantly facilitated by long word primes, $F_1(1, 57) = 2.61, p > .1$, and $F_2(1, 31) = 2.33, p > .1$. Responses to short target words were marginally faster following related short word primes than when following related long word primes, $F_1(1, 57) = 3.30, p < .1$, and $F_2(1, 31) = 3.70, p < .1$.

This pattern of greatest facilitation when prime syllables came from the same word as the target was confirmed by an analysis in which we used the difference between RTs following test and control primes as the dependent measure, with the length of the prime word and the length of the target word as independent variables (including version and item group factors as appropriate).

ANOVAs by participants and items showed that there was no overall difference in priming of short or long targets, $F_1 < 1$, and $F_2(1, 31) = 1.71, p > .1$, or from short or long primes ($F_1 < 1$ and $F_2 < 1$). Crucial in our findings, however, was a significant interaction between the prime length and target length factors, $F_1(1, 57) = 9.14, p < .01$, and $F_2(1, 31) = 8.76, p < .01$. This crossover interaction indicates that priming effects were greatest when the target word was identical to the prime stimulus.

We arcsine transformed error rates to stabilize variances (Winer, 1971), and we entered them into pairwise comparisons to assess differences in error rates following test and control primes (as carried out for the RT data). All of these comparisons failed to reach significance (all F_1 s < 1 and F_2 s < 1).

Discussion

The first important result that emerges from these analyses is that the measure of activation provided by cross-modal repetition priming does not show the bias toward short word interpretations that was observed in Experiment 1. Analysis of the magnitude of priming shows no overall difference between the priming of short and long target words. In this respect, therefore, the priming paradigm may provide a more transparent measure of lexical activation than the gating task.

A second important finding confirms the results obtained in Experiment 1; ambiguity between short and long word stimuli was not as pronounced as predicted by a phonemic analysis of our stimuli. The crossover interaction illustrated in Figure 3A shows that significantly greater priming was found for conditions in which the prime syllable came from the same word as the target. Because this difference was observed for prime stimuli in which participants heard only the first syllable of the test words (AP_1), some additional acoustic cue or cues biased lexical activation for embedded words and longer competitors in these two sets of test sequences.

In discussing lexical accounts of word segmentation (Grosjean, 1985; Luce, 1986; McQueen et al., 1995), researchers have argued that the ambiguity created by short words embedded in longer lexical items requires that recognition be delayed. However, the results reported thus far suggest that this ambiguity is not absolute and that other cues are available to the perceptual system to help distinguish short from long words—even at the offset of an embedded syllable. These additional cues modulate the hypothesized competition between short and long lexical items.

To investigate the ongoing competition between short and long lexical items, we reexamined another important finding from the gating study: evidence for competition between short and long words at AP_2 and beyond. The absence of a short word bias for the priming effects in Experiment 2A suggests that the activation of short and long words can be more accurately assessed using cross-modal priming. Therefore, we carried out three subsequent

³ On the basis of this procedure, we can calculate the maximum Type I error rate for each part of Experiment 2. The error rate for each family of comparisons is 0.1 (0.33×3); hence, for the two families of comparisons the experimental error rate is 0.20 (0.10×2). This is identical to the Type I error rate that results from a two-way repeated measures ANOVA for which four probability values are reported (two main effects, an interaction term, and a constant).

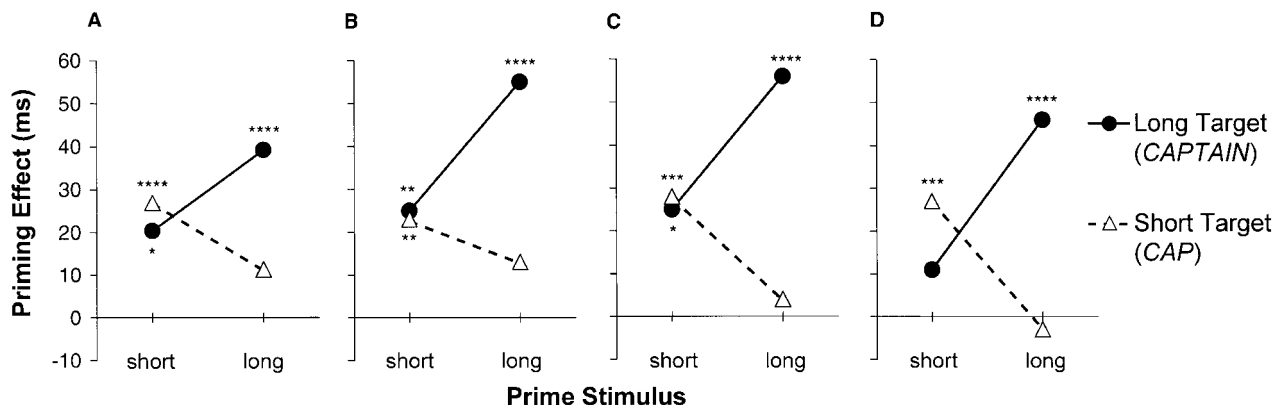


Figure 3. Experiment 2: Magnitude and significance of priming (difference between response times following test and control primes) from short (*cap tucked*) and long (*captain*) primes to short (*CAP*) and long (*CAPTAIN*) targets. A: Experiment 2A (AP₁). B: Experiment 2B (AP₂). C: Experiment 2C (AP₃). D: Experiment 2D (AP₃ + 100 ms). AP = alignment point. Asterisks indicate significant priming in comparisons of responses following test and control primes. * $p < .10$. ** $p < .05$. *** $p < .01$. **** $p < .001$.

experiments using this methodology, to examine the activation of short and long target words at later probe positions.

Experiments 2B, 2C, and 2D

We carried out three experiments similar to Experiment 2A, with the probe position changed, to track the activation of competing interpretations as an increasing amount of speech was presented. Because no significant difference between the two control primes was observed in Experiment 2A, the design of these experiments was changed so that only a single control prime was used for each target. This produced a six-condition design with three prime types (two test primes and one control prime) and two target types (short and long target words).

The three probe positions used for these experiments form a subset of the gates tested in Experiment 1. In Experiment 2B, the probe position was placed at AP₂ (after the onset segments of the syllable following the embedded syllable; e.g., after /t/ in *captain* and *cap tucked*). In the gating experiment, there was a marked increase in the proportion of long word responses at this probe position. The third alignment point (AP₃), which was used in Experiment 2C, was placed four pitch periods into the vowel of the second syllable, a position where the test stimuli diverged phonemically. Gating responses also diverged at this position, with reduced numbers of long word responses to short word stimuli. The final probe position used in Experiment 2D was 100 ms after AP₃ (Gate 7 in Experiment 1), a position at which the majority of gating responses correctly identified the target word for both short and long word stimuli. Therefore, we expected to find no facilitation for targets that did not match the prime words at this probe position.

Method

Participants. Across the three experiments, 181 paid participants from the Birkbeck CSL subject pool were tested, 54 on Experiment 2B, 72 on Experiment 2C, and 55 on Experiment 2D.⁴ None of the participants had taken part in any of the previous experiments.

Design and materials. Test materials were identical to those used in Experiment 2A and shown in Table 1. These stimuli were presented up to AP₂ in Experiment 2B, to AP₃ in Experiment 2C, and to 100 ms after AP₃ in Experiment 2D. The only significant divergence from Experiment 2A was in the number of control prime conditions used. A matched control prime was used for each target type (short control primes for short targets and long control primes for long targets), excluding the target conditions marked in Table 1.

We added related nonword fillers to ensure that form overlap between prime and target was not associated with a *yes* response; 20 filler items were used in Experiment 2B, and 27 nonword fillers were used in Experiments 2C and 2D. We added unrelated trials to reduce the relatedness proportion and to ensure an equal number of word and nonword targets. The overall proportion of experimental trials that contained a phonologically related word target was 18% in Experiment 2B and 14% in Experiments 2C and 2D.

Procedure. The procedure in Experiments 2B, 2C, and 2D was identical to that used previously, except for the change in probe position described above.

Results

RTs were analyzed following data trimming and carried out as before. Data from slow or error-prone participants were rejected (mean RT greater than 750 ms or errors greater than 12.5%). Responses to the target words *BRAN* and *BRANDY* were removed, along with individual data points over 1,200 ms. Mean RTs and error rates are shown in Table 3. Statistical analysis focused on the magnitude and significance of priming effects through pairwise comparisons and analysis of control minus test difference scores. As before, three pairwise comparisons were carried out for each target type using a corrected threshold of $p < .033$. These difference scores and the significance of priming effects in each experiment are shown in Figures 3B, 3C, and 3D.

⁴ Differences between these experiments in terms of the number of removed participants reflect differences in the amount of previous experience that groups of participants had with the lexical decision task.

Table 3
Mean Response Times (RTs; in Milliseconds) and Error Rates (in Percentages) by Prime and Target Type for Experiments 2B, 2C, and 2D

Prime type and word	Short target (CAP)		Long target (CAPTAIN)	
	RT	Error	RT	Error
Experiment 2B (AP ₂)				
Short test (<i>cap</i>)	546	2.2	582	4.2
Long test (<i>captain</i>)	556	2.6	552	1.5
Control (<i>palm</i> or <i>rifle</i>)	569	3.0	607	7.5
Experiment 2C (AP ₃)				
Short test (<i>cap</i>)	520	3.2	565	4.2
Long test (<i>captain</i>)	544	2.0	534	3.3
Control (<i>palm</i> or <i>rifle</i>)	548	3.1	590	8.3
Experiment 2D (AP ₃ + 100 ms)				
Short test (<i>cap</i>)	512	3.2	573	7.9
Long test (<i>captain</i>)	542	5.9	538	4.3
Control (<i>palm</i> or <i>rifle</i>)	539	5.7	584	6.8

Note. AP = alignment point.

Experiment 2B. The data of 5 of 54 participants were rejected because of slow and/or error-prone responses, and two data points over 1,200 ms were discarded (0.09% of the data). Results as shown in Table 3 and Figure 3B indicate consistent differences in the magnitude of priming, depending on the length of the words in the prime and target stimulus. This is despite the fact that both test prime stimuli would be identical in a phonemic analysis, a sequence matching the beginning of a long word (e.g., /kæpt/ matching *captain*).

Long word targets were significantly facilitated by long word primes, $F_1(1, 43) = 21.68, p < .001$, and $F_2(1, 33) = 24.71, p < .001$, with numerically weaker, though significant, priming of long targets by short word test stimuli, $F_1(1, 43) = 5.38, p < .05$, and $F_2(1, 33) = 5.30, p < .05$. Comparison of long word responses following short and long test primes showed that significantly greater priming was observed from long word stimuli, $F_1(1, 43) = 10.11, p < .01$, and $F_2(1, 33) = 10.73, p < .01$. Although long word interpretations were activated for the short word stimuli, greater priming was observed when the prime stimulus actually contained a long word. Conversely, responses to short targets were significantly facilitated by short word primes, $F_1(1, 43) = 6.59, p < .05$, and $F_2(1, 33) = 5.88, p < .05$, though not by long word primes, $F_1(1, 43) = 2.08, p > .1$, and $F_2(1, 33) = 2.52, p > .1$. For short word targets, however, the difference between responses following short and long test primes was nonsignificant, $F_1(1, 43) = 1.19, p > .1$, and $F_2(1, 33) = 1.28, p > .1$. This pattern suggests that differences in the activation of short words by long and short word stimuli were less marked at this probe position, possibly as a consequence of increased activation of long word interpretations.

Differences in the interpretation of short and long word stimuli are also illustrated by ANOVAs we carried out on test-control difference scores, with the factors prime and target length. As at

the previous probe position, this analysis showed an interaction between prime and target length, $F_1(1, 43) = 8.72, p < .01$, and $F_2(1, 33) = 9.43, p < .01$; however, there were main effects of prime length, $F_1(1, 43) = 4.27, p < .05$, and $F_2(1, 33) = 2.92, p < .1$, and target length, $F_1(1, 43) = 4.22, p < .05$, and $F_2(1, 33) = 3.64, p < .1$, although these effects were only marginally significant by items. This pattern confirms the impression given in Figure 3B, namely that at this probe position, short word stimuli remained ambiguous—with both short and long word interpretations active.

Pairwise comparisons of error rates showed significant facilitation of responses to long words. Participants made significantly fewer errors to long targets when they were preceded by a long word prime, $F_1(1, 43) = 19.66, p < .001$, and $F_2(1, 33) = 9.92, p < .01$, compared with error rates following control primes. There was also a marginal reduction in error rate when long word targets followed a short word prime, $F_1(1, 43) = 2.49, p > .1$, and $F_2(1, 33) = 3.56, p < .1$, with a marginally significant difference between error rates following short and long test primes, $F_1(1, 43) = 3.05, p > .1$, and $F_2(1, 33) = 4.14, p < .1$. There were no significant differences in error rates to monosyllabic targets following short or long test primes or control primes (all $ps > .1$).

Experiment 2C. Out of 72 participants tested, data from 14 participants were discarded for slow or error-prone responses. An additional participant whose mean RT was more than two standard deviations faster than any other participant was also removed. Also excluded were 17 individual outlying responses slower than 1,200 ms (0.76% of the data). Analysis of priming effects as shown in Figure 3C indicates a similar pattern to that observed at the previous probe position, with long word stimuli showing very little ambiguity and short word stimuli remaining ambiguous. This is despite the presence of segmental information in the vowel of the second syllable that helped participants to discriminate short and long word sequences in gating (e.g., /kæpt/ or /kæpt/ in *cap tucked* or *captain*, respectively).

Pairwise comparisons showed that responses to long target words were significantly faster following long test primes than when following control primes, $F_1(1, 51) = 45.37, p < .001$, and $F_2(1, 33) = 26.12, p < .001$, or when following short test primes, $F_1(1, 51) = 8.71, p < .01$, and $F_2(1, 33) = 13.58, p < .01$. Facilitation of long word targets by short test primes compared with controls was marginally significant at this probe position, $F_1(1, 51) = 6.84, p < .05$, and $F_2(1, 33) = 3.92, p < .1$. Conversely, responses to short targets were facilitated by short word stimuli compared with control primes, $F_1(1, 51) = 15.05, p < .001$, and $F_2(1, 33) = 9.57, p < .01$, and compared with long test primes, $F_1(1, 51) = 9.83, p < .01$, and $F_2(1, 33) = 8.58, p < .01$, whereas responses to short words following long test primes did not differ from control primes ($F_1 < 1$ and $F_2 < 1$). Thus, although short word stimuli produced priming for both short and long targets, significant priming was only observed for long targets from long word stimuli.

The ambiguity of short word stimuli and the comparative lack of ambiguity of long word stimuli was also suggested by analysis of test-control difference scores. ANOVAs showed a significant interaction between prime and target type, $F_1(1, 51) = 15.78, p < .001$, and $F_2(1, 33) = 20.92, p < .001$, such that greatest facilitation was observed only when prime and targets matched. Contrary to Experiment 2A, there was also a main effect of target type,

$F_1(1, 51) = 5.87, p < .05$, and $F_2(1, 33) = 3.18, p < .1$, with greater overall priming for long targets. The main effect of prime type was nonsignificant ($F_1 < 1$ and $F_2 < 1$).

Facilitation for long targets was also observed in comparisons of error rates following test and control primes. Participants made significantly fewer errors to long test words after hearing long test primes than after hearing control primes, $F_1(1, 51) = 7.92, p < 0.01$, and $F_2(1, 33) = 10.91, p < 0.01$. Error rates for long targets were reduced following short word primes compared with following controls, $F_1(1, 51) = 4.84, p < 0.05$, and $F_2(1, 33) = 4.21, p < 0.05$, and there was no difference between error rates for long words following short and long test primes ($F_1 < 1$ and $F_2 < 1$). No significant differences in error rates were found for short word targets (all $ps > .1$)

Results obtained in this experiment show that long word interpretations were still active for short word stimuli. This finding might be unexpected on the basis of gating data from Experiment 1, in which the majority of responses favored the short word at this probe position. Given that the prime stimuli differed phonemically at AP₃ (in the vowel of the second syllable), we had expected a clear preference for short word interpretations of short word stimuli to emerge at this probe position. This discrepancy between the results obtained in gating and those obtained in cross-modal priming may simply reflect the greater time available for the processing of stimuli in the gating task. However, we needed to rule out the possibility that there was some systematic difference in the measures of lexical activation obtained for short and long words in the cross-modal priming experiments. Hence, we carried out the further priming experiment (2D) at a later probe position.

Experiment 2D. Out of 55 participants tested, the data of 8 were rejected because of slow and/or error-prone responses. Also removed were four data points over 1,200 ms (0.22% of the data). The pattern of priming effects shown in Figure 3D suggests that both short and long word stimuli were unambiguous at this position, as would be expected given the majority of correct responses at this probe position in Experiment 1 (Gate 7 in Figure 2).

Planned comparisons showed that responses to long targets were facilitated by long test stimuli compared with control primes, $F_1(1, 41) = 21.11, p < .001$, and $F_2(1, 33) = 19.57, p < .001$, and compared with short test primes, $F_1(1, 41) = 17.24, p < .001$, and $F_2(1, 33) = 16.83, p < .001$. In contrast to the previous three probe positions, long words were not significantly facilitated by short test stimuli compared with controls, $F_1(1, 41) = 1.13, p > .1$, and $F_2(1, 33) = 1.83, p > .1$. Responses to short targets were significantly facilitated following short test primes compared with following controls, $F_1(1, 41) = 17.06, p < .001$, and $F_2(1, 33) = 9.28, p < .01$, and compared with following long test primes, $F_1(1, 41) = 10.63, p < .01$, and $F_2(1, 33) = 10.50, p < .01$. Short word responses following long test primes did not differ significantly from responses following control primes ($F_1 < 1$ and $F_2 < 1$).

These results suggest that neither short nor long word stimuli are ambiguous at this probe position. ANOVAs using the difference between RTs following test and control primes as the dependant variable failed to show main effects of either prime length ($F_1 < 1$ and $F_2 < 1$) or target length, $F_1(1, 41) = 1.32, p > .1$, and $F_2(1, 33) = 1.63, p > .1$, yet there was a highly significant interaction between these factors, $F_1(1, 41) = 27.59, p < .001$, and $F_2(1, 33) = 30.08, p < .001$. This crossover interaction in the magnitude

of priming by prime and target length demonstrates increased facilitation for targets that were identical to the test prime.

The reduced facilitation for related but nonidentical primes is also shown in error rate data. Error rates for long word targets were marginally lower following long test primes than following short test primes, $F_1(1, 41) = 4.09, p < .1$, and $F_2(1, 33) = 4.43, p < .1$. Furthermore, marginally fewer lexical decision errors were made to short targets following short test primes compared with following long test primes, $F_1(1, 41) = 3.51, p < .1$, and $F_2(1, 33) = 2.19, p > .1$. All other comparisons of error rates were nonsignificant (all $ps > .1$).

Combined analysis. We analyzed the data from all four parts of Experiment 2 to determine whether there was any change in the pattern of priming across the four probe positions tested. To perform this between-experiments comparison, we transformed RTs for each participant into z scores (i.e., RTs for target words were expressed as a number of standard deviations greater or less than each participant's mean), removing differences between participants in the speed or variability of their lexical decision responses (cf. Zwitserlood, 1989; Moss, McCormick, & Tyler, 1997). Differences between RTs following test and control primes were entered into a three-way ANOVA with the factors of prime length, target length, and probe position. Probe position was coded as a within-groups factor in the analyses by items and as a between-groups factor in the analysis by participants.

This combined analysis failed to show any main effect of probe position, $F_1 < 1$, and $F_2(3, 114) = 1.10, p > .1$, or any interaction between probe position and other factors (Probe Position \times Target Length and Probe Position \times Prime Length; all $F_s < 1$). Thus, despite changes in the pattern observed at each probe position, there was no simple change in priming effects obtained in each of the four parts of Experiment 2. This combined analysis showed a main effect of target type, $F_1(1, 214) = 8.02, p < .01$, and $F_2(1, 38) = 7.78, p < .01$, indicating greater overall priming of long targets across the four experiments; a possible explanation, which we explore later, is that short word stimuli in lexical garden-path sequences increase the activation of long word interpretations. There was no main effect of prime type, $F_1(1, 214) = 2.46, p > .1$, and $F_2(1, 38) = 2.07, p > .1$; any disruption that may have been caused by lexical garden-path sequences did not reduce the overall level of priming observed for short word stimuli. This analysis did, however, reveal a significant interaction between prime and target length, $F_1(1, 214) = 57.81, p < .001$, and $F_2(1, 38) = 57.27, p < .001$, equivalent to the effect shown in each of the four experiments. Priming effects were stronger between identical prime and targets (see Figure 3) than between phonologically related primes and targets. This Prime \times Target interaction was, however, unaffected by probe position, $F_1(3, 214) = 1.21, p > .1$, and $F_2(3, 114) = 2.07, p > .1$, indicating that information was present at all four probe positions to distinguish short from long stimuli—even at early probe positions for which both short and long word primes were phonemically identical.

Discussion

The results of Experiment 2 support one of the main findings from Experiment 1: Information is available early in the processing of embedded word stimuli to assist the perceptual system in distinguishing between short words and longer competitors. The

significant interaction in the magnitude of priming between prime and target length in all four experiments indicates that differences in the acoustic form of short and long word stimuli directly affect levels of lexical activation for short and long word interpretations. One consequence of these acoustic cues is that in all four parts of Experiment 2, greater priming of long targets was observed from long word primes than from short words that match the start of long words. Similarly, at two out of four probe positions tested, the converse was observed for short word targets, with greater priming by short word sequences than by long word sequences containing the short word as an embedding. This finding indicates that short words that are embedded at the start of longer lexical items do not produce the severe ambiguity that would be predicted by a phonemic analysis.

Despite the effectiveness of early cues that allowed participants to distinguish between short words and the longer words in which they were embedded, we also see evidence that information coming after the offset of the embedded syllable affected the processing of lexical garden-path stimuli. Facilitation of long targets was observed for short word stimuli in Experiments 2B and 2C, indicating that longer interpretations remained active until later in the processing of garden-path stimuli. Because long word interpretations remained active, it is likely that competition from these longer words produced the delayed recognition of onset-embedded words observed in the gating study. One difference, however, between results in gating and cross-modal priming is that activation of long word interpretations of short word sequences appeared to decline at an earlier point in the speech stream in the gating experiment compared with cross-modal priming. It may be that the additional processing time that was available in gating (due to off-line responses and repeated presentations) allowed more use to be made of the mismatching vowel information that was present at AP₃. For cases in which processing time was reduced (as in the repetition priming experiments), effects of mismatch could have been reduced (cf. Zwitserlood & Schriefers, 1995). It is only at the final probe position tested in cross-modal priming (see Figure 3D) that short word stimuli were as unambiguous as long word stimuli.

Activation of longer competitors of short word stimuli has also been observed in other cross-modal priming studies using short embedded words as primes. For instance, Tabossi et al. (1995) reported that lexical items formed by combining speech across word boundaries are activated in connected speech (e.g., words related to *visite* [visit] are primed by the sequence *visi tediati* [faces bored]). Similarly, Gow and Gordon (1995) observed priming for targets related to single-word interpretations of two-word sequences (e.g., priming of the target *FLOWER* from the sequence *two lips*, as well as from *tulips*). However, one difference from previous experiments is that we consistently observed significantly greater priming of long word targets from sequences that contained the long word than from related short word stimuli. This difference in the magnitude of priming was even observed when the fragments of short and long word sequences contained identical phonemes (e.g., at AP₁); thus, the ambiguity of onset-embedded words is not as severe as was suggested by the results of Tabossi et al. or Gow and Gordon.

Our results demonstrate an absence of the extreme ambiguity predicted by accounts of spoken word recognition in which lexical access is driven by a phonemically categorized input. However, it was incumbent upon us to demonstrate that this finding does not

reflect a peculiarity of our experimental stimuli but could be generalized to other speakers' productions of these same items. Despite our best efforts of to produce the paired sentences without unnatural or inappropriate emphasis on the target words, it is conceivable that the relevant acoustic cues could be less marked in recordings made by speakers naive to the nature of the stimulus materials. Although the acoustic cues that we described are consistent with those documented in the acoustic-phonetics literature (e.g., Klatt, 1976; Lehiste, 1972), it was valuable to carry out a more rigorous investigation of the acoustic and perceptual properties of syllables in embedded words, because between-speaker variability has been reported in previous studies (Barry, 1981)

Acoustic Properties of Embedded Word Stimuli

To determine whether the acoustic cues identified in the original recordings could be generalized to other speakers, we made three additional sets of recordings of the stimulus items. Three speakers produced these sentences (2 male and 1 female), all of whom were naive to the nature of the stimulus sentences. They were each instructed to read a (different) randomly ordered list of the short and long word test sentences. In all three lists, at least 20 sentences intervened between pairs of test items. We expected that this spacing would minimize the potential for contrastive cues to be included in these pairs of stimulus materials. Three alignment points were marked in these recordings and the same acoustic analyses were carried out as reported for the original test stimuli.

These acoustic analyses confirmed many of the differences between short and long word sequences that were observed in the original test recordings. The most salient difference was that embedded syllables (e.g., /kæp/ in *cap* and *captain*) were of significantly greater duration in short words than in a longer word (paired *t* tests were significant at $p < .001$ in each set of stimuli). Differences in the duration of the voiced portion of the syllable were equally significant (all $ps < .001$), suggesting that vowel duration accounted for the majority of the difference in syllable duration. Again, syllables in short words had a lower mean fundamental frequency than the equivalent syllable in a longer word, though this difference was more marked than previously ($p < .001$ in all three sets of recordings; in the original test stimuli, this difference was of only marginal significance). The presence of prosodic boundaries between short words and continuations in these new recordings could have introduced a decline in F_0 that accounts for this increased difference in mean fundamental frequency. For two out of three sets of recordings (as for the original test set), there were no significant differences in the mean RMS energy of embedded syllables from short and long words ($p > .1$), although one set of recordings did show significantly greater acoustic energy in syllables from short words ($p < .05$). These analyses show that the three sets of naive recordings shared many of the same acoustic differences that were present in the original test stimuli. Acoustic differences between syllables in short and long words were, if anything, of greater magnitude in naive-speaker recordings than in the original test set.

Further acoustic analysis focused on the onset of the following syllable (e.g., /t/ in *tucked* or *captain*). Unlike the original set of test recordings, these segments were of greater duration in short word stimuli than in long word stimuli (i.e., segments were longer when they were word initial rather than word medial). This finding

was significant for all three sets of recordings ($p < .001$). This difference illustrates the care that was taken in recording the original test set to not include prosodic boundaries between short and long words. Prosodic boundaries between short words and continuations (e.g., between *cap* and *tucked*) would increase the duration of the initial segment of the following word (Klatt, 1976). This finding therefore indicates that our original test recordings were likely to be more ambiguous than stimuli produced by naive speakers.

Given that there were differences between our test recordings and those produced by speakers who were naive to the nature of the experimental stimuli, it was important to assess the contribution of this acoustic variability to the perception of embedded words in connected speech. Because acoustic differences between short and long word stimuli were more marked in these recordings, we might expect that the discrimination of syllables in short and long words would be correspondingly enhanced. Therefore, we re-examined a crucial finding from Experiments 1 and 2—namely, that responses to short and long word stimuli differed at the offset of the embedded syllable. We compared results obtained using the original test recordings with short and long stimuli produced by naive speakers.

Experiment 3: Forced-Choice Experiment

Both gating and priming data showed significant differences between responses to short and long word stimuli at the earliest alignment point tested. At the offset of the embedded syllable, lexical activations were biased toward appropriate interpretations of short and long word sequences. Comparisons with the additional recordings showed that the test stimuli used for Experiments 1 and 2 contained a weaker set of acoustic differences between short and long words than were found in naive recordings. Therefore, it was important to extend our experimental results by comparison with naively produced stimuli. We used a two-alternative, forced-choice task (similar to that used in the forced-choice cloze pretest and by Mattys, 2000) to investigate participants' interpretations of stimuli presented up to the offset of the embedded syllable (AP_1). A 6-point rating scale was used so that a single response indicated overall preference and confidence ratings.

Method

Participants. Forty participants from the Centre for Speech and Language, Cambridge, subject pool were tested. Most were Cambridge University students. All were aged between 18 and 45 years, native speakers of British English, with normal hearing and no history of language impairment. They were paid for their participation; none had taken part in any of the previous experiments.

Design and materials. Participants were presented with fragments of test sentences presented up to the offset of an embedded syllable (AP_1) from either a short or a long word. They made responses on a 6-point scale, indicating which of a pair of target words the syllable came from as well as their confidence in their decision. The four sets of 40 sentence pairs (the original test set and three sets recorded by naive speakers) were pseudo-randomly divided into two lists (with a different randomization for each set of recordings), such that each list contained one of each sentence pair. Sentences from a single speaker were presented in a block, with items randomly ordered for each participant. Each participant was randomly assigned to one of eight test versions (two different lists presented in four different orderings of the test blocks).

Procedure. Participants were tested in groups of 4, in booths in a quiet room. Speech files were played over closed-ear headphones from a Pentium PC running DMDX software (Forster & Forster, 1999). The computer screen displayed a 6-point scale with the two target words presented below either end of the scale (short and long words were pseudorandomly assigned to be either 1 or 6 on the scale). Middle ratings on the scale were labeled as *guess*, and extreme ratings were labeled *confident*. At the offset of the embedded syllable, the speech was cut off and participants were prompted for a numerical response on the computer keyboard. Following each response, the screen was cleared and the next item presented after a short delay. Participants were provided with 10 practice items and a short break between each of the four blocks, in a testing session lasting approximately 20 min.

Results

Participants' responses were transformed so that short word responses were coded as 1 to 3, and averaged over participants and items for short and long stimuli from each set of recordings. Responses were also categorized as indicating an overall preference for a short or a long word, and a signal detection measure of sensitivity (d') was derived from these response proportions (Macmillan & Creelman, 1991). These measures are summarized for each of the four sets of stimuli in Table 4.

For the original set of recordings, a repeated measures ANOVA showed a highly significant difference in ratings for short and long word stimuli, $F_1(1, 39) = 122.83$, $p < .001$, and $F_2(1, 39) = 38.37$, $p < .001$. The same effect was also observed in an analysis of the proportion of responses matching a short and a long word, $F_1(1, 39) = 143.86$, $p < .001$, and $F_2(1, 39) = 45.49$, $p < .001$. Differences between ratings and response proportions were also highly significant for all three sets of naive recordings (all $ps < .001$). Bonferroni-corrected post hoc comparisons were carried out to assess differences in sensitivity (d') between the four sets of stimuli. These comparisons showed that all three sets of naive recordings were more easily discriminated than the original test recordings (for differences between original recordings and naive recordings, all $ps < .01$), whereas the naive recordings did not differ from one another ($p > .1$).

Correlations between the magnitude of acoustic differences between short and long word stimuli (percentage of changes in duration, F_0 , and RMS energy) and sensitivity (d') showed a significant correlation with the magnitude of duration differences

Table 4
Results of Experiment 3: Two-Alternative Forced-Choice Gating Test at AP_1

Stimulus set	d'	Proportion of long word responses		Mean rating ^a	
		Short stimuli	Long stimuli	Short stimuli	Long stimuli
Test items	1.01	.38	.69	3.01	4.21
Recording 1	1.94	.19	.78	2.18	4.57
Recording 2	1.76	.32	.83	2.72	4.77
Recording 3	1.72	.28	.80	2.52	4.61

Note. Differences between responses to short and long stimuli were all significant at $p < .001$. AP = alignment point.

^a 1 = short word; 3.5 = neutral; 6 = long word.

for each stimulus pair, $r(160) = .504$, $p < .001$. The correlation between changes in mean F_0 and sensitivity was also significant, $r(160) = .423$, $p < .001$. Differences in RMS amplitude, however, did not significantly correlate with sensitivity, $r(160) = -.045$, $p > .1$.

Discussion

These findings confirm that acoustic cues are present in embedded syllables that help to distinguish between short and long words. Far from being a peculiarity of the original test recordings, these cues were even more effective in recordings made by experimentally naive speakers. Because the original test recordings excluded prosodic boundaries between short words and their continuations, it is possible that they omitted a potentially useful cue to the location of a word boundary. Nonetheless, participants still showed significant differences between responses to short and long word stimuli for the original test set in the absence of prosodic boundary cues.

Acoustic cues to word length or word boundaries. The results of Experiments 1–3 indicate that the activation of short and long lexical items was biased toward the appropriate interpretation of sequences containing embedded words and longer competitors. Two types of acoustic cues have been described in the literature that might account for the discrimination of short and long word sequences: (a) qualitative changes in the initial segments of words compared with segments that are in the middle or at the offset of words (Gow & Gordon, 1995; Lehiste, 1960; Nakatani & Dukes, 1977), and (b) duration differences between segments and syllables in monosyllables and longer words (Klatt, 1976; Lehiste, 1972; Nakatani & Schaffer, 1978). Caution is required in drawing strong conclusions that our results implicate one or another type of cue. The stimuli used in our study were not directly manipulated to include only certain acoustic differences. However, because controlled fragments of speech were presented to participants, we could at least determine what portion of the speech signal contained the relevant acoustic differences. Because the participants in our experiments were able to distinguish long words from embedded words when only an embedded syllable was presented, we suggest that cues in the onset of the following word were unlikely to be responsible for the pattern of results that we obtained. In addition, differences in the duration of word onsets were not statistically reliable for the main set of test stimuli used in our experiments.

Therefore, we conclude that acoustic differences between the initial syllable of a bisyllable and the same syllable as a monosyllabic word are the cue that is most likely to be responsible for differences in the activation of short words and longer competitors. Two measured acoustic properties were shown to differ reliably between syllables in short and long words—mean syllable duration and mean fundamental frequency. In Experiment 3, the magnitude of these duration and F_0 differences was correlated with listeners' sensitivity to differences between syllables from short and long words. However, these correlations are not sufficient to demonstrate that these cues (as opposed to some more subtle acoustic difference) are responsible for listeners' ability to distinguish syllables from short and long words. Follow-up experiments that directly manipulate duration and fundamental frequency will be

required to establish that these cues alone are sufficient for the discrimination of syllables in short and long words.

Delayed identification of embedded words. Despite the lack of absolute ambiguity between monosyllabic words and the initial syllable of longer competitors, we still have evidence that the identification of short word stimuli is delayed with respect to word offset. For instance, in gating, short words were not correctly identified until later gates, when long word interpretations could be ruled out. Similarly, increased activation for long lexical items was observed in Experiment 2, because significant priming was observed for long targets at probe positions after the offset of the embedded word (AP₂ and AP₃). These results are therefore consistent with accounts in which competition between embedded words and longer lexical items reduces the activation of short word stimuli. Similar results have also been obtained by McQueen et al. (1994), who showed that the detection of monosyllables is more difficult in sequences that can be continued to form a word (e.g., *sack* is spotted less reliably in /sækræf/ from *sacrifice*, compared with the nonword sequence /sækræk/).

However, it is unclear whether competition from longer lexical items is the only explanation for this effect. It may be that reduced facilitation for short targets in Experiment 2 reflects an intrinsic difference in priming effects for short words rather than specific competition from longer lexical items. For instance, short words may be more weakly or more slowly activated during identification as a consequence of their competitor environment—rather than as a consequence of the coactivation of longer words for these sequences. Alternatively, the repetition priming paradigm may be less sensitive for detecting facilitation of short words because of faster lexical decision responses to short visual targets (see Forster & Davis, 1991, for similar findings in masked visual priming).

Therefore, it was valuable to examine the identification of onset-embedded words in sequences that would not be expected to activate longer lexical items as strongly. In two follow-up experiments, we investigated the identification of short words in sequences in which the continuations of short word stimuli did not temporarily match longer words. Listeners' interpretations of these non-garden-path sequences were tested using gating and cross-modal priming in Experiments 4 and 5.

Experiments 4 and 5: Non-Garden-Path Stimuli

The results of Experiments 1 and 2 suggest that the identification of short words may be disrupted by lexical garden paths (continuations that match a longer word). Therefore, the goal of Experiments 4 and 5 was to investigate the time course of recognition of onset-embedded words under conditions in which we predicted less disruption from longer competitors. We used sequences containing a short word with a continuation that had an immediate mismatch with all the longer lexical items in which the short word was embedded (such as the sequence *cap looking*, for which the short word combined with the onset of the following syllable, /kæpl/, does not match any longer word). To allow comparisons with the previous experiments, we used the same off-line (gating) and online (repetition priming) methods with the same pairs of short and long words, altering only the words after the short target word. Because the critical comparisons involved the effectiveness of short word primes, we did not need to test the

long word materials a second time. These experiments therefore used a single set of short-word test stimuli.

We used the same set of 40 onset-embedded monosyllables as used previously. The short words were placed in the same sentential context as before, but with the words following the test word changed so that the onset of the following word mismatched with all longer lexical items. For instance, for the word *cap*, other words that start with the syllable [kæp] include: *captain*, *caption*, *capsule*, and *capture*. Thus, in picking a continuation that begins with the phoneme /l/, as in the sequence *cap looking*, the embedded word plus continuation would mismatch with all longer lexical items.⁵ Because phoneme sequences that do not occur word internally might provide a prelexical cue to a word boundary, we chose continuations such that the segments on either side of the word boundary were found word internally in other English words; that is, these sequences were phonotactically legal (Mattys et al., 1999). For the example *cap*, other words in CELEX with the segments /æp/ at the end of a syllable include *clapboard*, *haphazard*, *napkin*, *chaplain*, and *entrapment*. By analogy, with the word *chaplain*, continuations starting with the segment /l/ would not provide a phonotactic cue to a word boundary. Hence, the test sentence for the monosyllable *cap* was *The soldier saluted the flag with his cap looking slightly crumpled*. Each of the onset-embedded monosyllables used in Experiments 1–3 was placed in a non-garden-path sequence. The complete set of 40 sentences is shown in the Appendix.

These materials were recorded by Matthew H. Davis, with care being taken to ensure that no prosodic boundary was placed after the offset of the target word. Three alignment points, equivalent to those used previously, were marked for these stimuli (AP₁ at the offset of the first syllable of the test word, AP₂ after the onset segments of the following syllable, and AP₃ in the vowel of the second syllable). Durations of these sections did not significantly differ from the short word stimuli that included lexical garden paths.

Experiment 4: Gating

We used the gating task to investigate whether the identification of embedded monosyllables with continuations that mismatch with longer competitors follows a different pattern from that observed in Experiment 1. This investigation allowed us to determine whether the delayed recognition observed in the earlier gating experiment reflects competition from longer lexical items.

Method

Participants. Eleven paid participants from the Birkbeck CSL subject pool took part in the experiment. None had taken part in any of the previous experiments.

Design, materials, and procedure. Design and procedure followed Experiment 1, except that only a single set of test stimuli were used containing short words with non-garden-path following contexts. Gates were set up as before: at the three alignment points, and at two additional gates before AP₁ and five gates after AP₃. The gated stimuli were presented in successive fragments in a single test version using all 40 test stimuli accompanied by 16 filler items.

Results and Discussion

Results were analyzed in terms of the proportion of responses matching either of the target words. As in Experiment 1, three

items (*ban*, *bran*, and *win*) produced a disproportionate number of errors, and data from these items was discarded. The proportions of responses matching either short or long target words at different gates are shown in Figure 4. Also included in this graph are data from the short word stimuli of Experiment 1.

Participants did not produce significant numbers of long word responses at gates where the following word could be heard. This change from Experiment 1 is especially apparent at AP₂, where information about the onset of the following word becomes available. However, there is also a discrepancy at AP₁, with significantly more short word responses to non-garden-path stimuli, $F_1(1, 31) = 24.10, p < .001$, and $F_2(1, 36) = 18.91, p < .001$, and significantly fewer long word responses, $F_1(1, 31) = 9.48, p < .01$, and $F_2(1, 36) = 5.30, p < .05$. This finding suggests a strong influence of coarticulated information on these results and those obtained for garden-path stimuli. Although the initial alignment point was placed at the offset of the embedded word, we expected that some information from the onset of the following word would be available at this probe position. Differences between lexical garden-path and non-garden-path stimuli were therefore apparent before the sections marked as containing that particular phoneme.

The effect of lexical garden-path continuations could also be seen in the overall proportion of correct responses for the two types of stimuli. In this experiment, all participants correctly identified the target words at the final gate. For garden-path stimuli in Experiment 1, only 94% correct responses were given for embedded words at Gate 10—significantly fewer than for non-garden-path stimuli, $F_1(1, 31) = 10.03, p < .01$, and $F_2(1, 36) = 20.96, p < .001$. Furthermore, correct responses for non-garden-path stimuli were isolated before the offset of the embedded word. The mean isolation point for non-garden-path stimuli (the point at which participants produced the correct response without changes at subsequent gates; Grosjean, 1980, 1996) was 268 ms. A paired *t* test showed this isolation point to be significantly before the offset of the embedded word (303 ms), $t(36) = 3.29, p < .01$.

This early isolation point indicates that responses to short word stimuli in non-garden-path contexts remained consistent before and after their acoustic offset, contrary to Experiment 1. It is possible that other gating experiments that demonstrated post-offset isolation points for words in connected speech (Bard et al., 1988; Grosjean, 1985) contained lexical garden paths equivalent to those that were deliberately constructed in the materials for Experiments 1–3. Alternatively, it may be that our stimuli provided a stronger contextual constraint than those used in previous experiments, allowing earlier isolation in the gating task (Tyler & Wesels, 1983). Either way, it appears that the identification of short embedded words was easier when these words were presented with continuations that immediately mismatch with longer lexical items—consistent with the predictions of lexical competition accounts.

⁵ Although it is possible that novel morphological variants (such as *capless*) would match these stimuli, we ensured that no words listed in the CELEX database matched the embedded word and the onset of the following syllable.

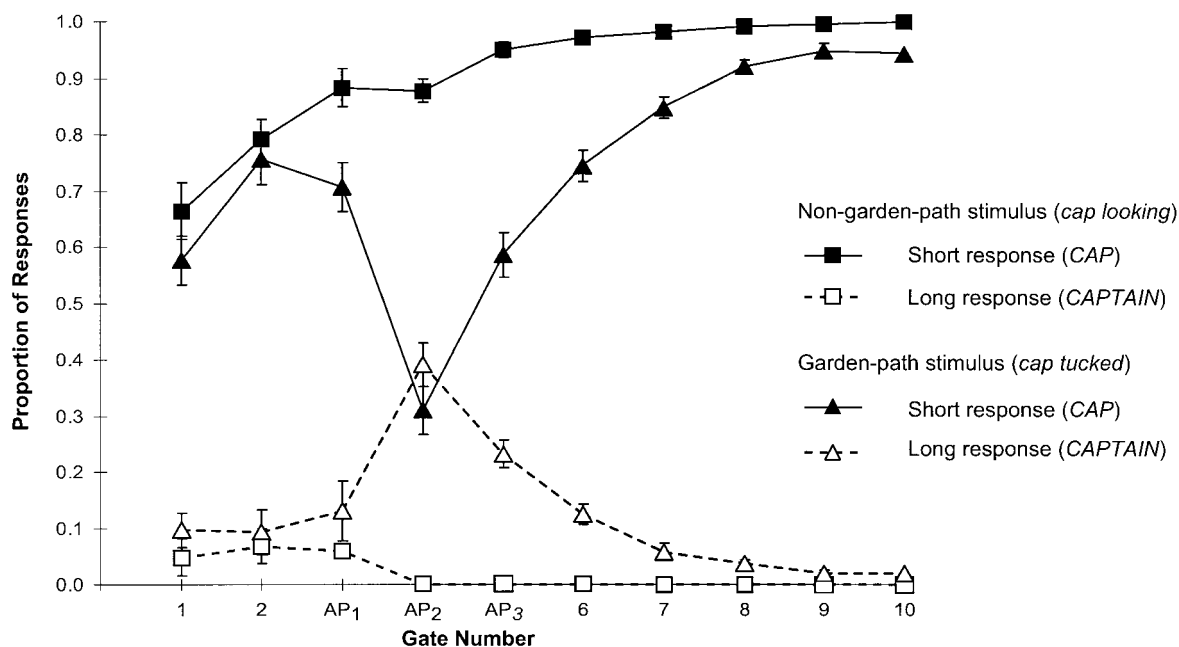


Figure 4. Experiment 4: Proportions of responses at each gate matching short (*CAP*) and long (*CAPTAIN*) target words for short stimuli with garden-path (*cap tucked*) or non-garden-path (*cap looking*) following contexts. Error bars show one standard error. Data for garden-path sequences are from Experiment 1. AP = alignment point.

Experiments 5A and 5B: Cross-Modal Priming

In the main further study, we used cross-modal repetition priming to provide an online measure of the activation of onset-embedded words in non-garden-path contexts. We compared priming of short and long target words and longer competitors to establish whether these stimuli activate competing interpretations. As in Experiment 4, we investigated a single set of test stimuli containing short words in non-garden-path contexts. Measuring the priming of short and long targets required four experimental conditions (two prime types and two target types) at each probe position. To reduce the number of experimental versions required, we combined control prime conditions across two probe positions, such that a total of 12 experimental conditions were required (four test prime positions and two control primes combined with two target types). These 12 experimental conditions were tested in two 6-version experiments.

Method

Participants. We tested 114 participants from the Birkbeck CSL subject pool (56 on Experiment 5A and 58 on Experiment 5B). A shortage of previously untested participants meant that approximately 20 of the participants had taken part in the repetition priming experiments in Experiment 2A. However, none had been tested within the previous 12 months and none had heard the test stimuli used in this experiment.

Design and procedure. The 40 test stimuli were paired with the same control prime sentences used in Experiment 2. Two different probe positions were tested with a single control prime, such that all four probe positions were tested in two 6-version experiments. Because the same number of data points per condition were tested in a single version, these experiments would have the same power to detect significant priming as

did Experiments 2B, 2C, and 2D. Experiment 5A tested priming at AP₁ and 100 ms after AP₃, and Experiment 5B probed at AP₂ and AP₃. Control primes in each experiment were presented up to a point equivalent to the earlier of the two probe positions.

In all other respects, both experiments were identical in design and procedure to Experiment 2: Related nonword fillers were added to ensure that an equal proportion of nonword targets were related to the auditory prime; unrelated word and nonword fillers were added to ensure that 50% of targets were words. Counting the 20 practice items and 10 dummy items, 14% of trials contained a word target preceded by a related test prime.

Results and Discussion

RT data were analyzed following the exclusion of data from slow or error-prone participants (mean RT greater than 750 ms or error rates greater than 12.5% on test words); the data from 8 participants from Experiment 5A and 11 participants from Experiment 5B were removed. The target pair *BRAN* and *BRANDY* was excluded, along with outlying responses slower than 1,200 ms: four data points (0.21%) from Experiment 5A and one data point (0.054%) from Experiment 5B. RTs and error rates for each prime and target type are shown in Table 5. We used the modified Bonferroni correction for planned comparisons in our statistical analysis, to evaluate the magnitude and significance of differences between RTs following test and control primes. In analyses comparing the priming effects found in different experiments, we used differences between z -score transformed control and test prime RTs as the dependent measure.

Experiment 5A. Responses to short word targets were significantly facilitated by test primes at AP₁, $F_1(1, 42) = 5.87$, $p < .05$, and $F_2(1, 33) = 5.00$, $p < .05$, and 100 ms after AP₃, $F_1(1, 42) =$

Table 5
Mean Response Times (RTs; in Milliseconds) and Error Rates
(in Percentages) by Prime and Target Type for
Experiments 5A and 5B

Prime type and probe position	Short target (CAP)		Long target (CAPTAIN)	
	RT	Error	RT	Error
Experiment 5A				
Test (AP ₁)	488	2.7	558	5.5
Test (AP ₃ + 100 ms)	480	1.6	549	7.6
Control	508	4.5	560	7.5
Experiment 5B				
Test (AP ₂)	499	1.0	556	5.7
Test (AP ₃)	500	3.5	569	5.9
Control	520	5.7	551	6.2

Note. AP = alignment point.

14.73, $p < .001$, and $F_2(1, 33) = 13.62$, $p < .001$. There was no significant difference in the priming of short word targets at these two probe positions, $F_1(1, 42) = 1.54$, $p > .1$, and $F_2(1, 33) = 1.21$, $p > .1$. No significant priming effects were observed for long targets (all F_1 s < 1 and all F_2 s < 1) nor any difference between the two test prime conditions ($F_1 < 1$ and $F_2 < 1$). Differences in error rates following test and control primes also failed to show any significant differences for short or long words (all $ps > .1$); however, there was a significant reduction in error rates for short words following test primes compared with control primes at 100 ms after AP₃, though this was nonsignificant by items, $F_1(1, 42) = 4.98$, $p < .05$, and $F_2(1, 33) = 2.35$, $p > .1$. The magnitude and significance of the priming effects observed for each prime and target type are shown in Figure 5.

Experiment 5B. Pairwise comparisons shown in Figure 5 indicated marginally significant priming for short word targets at AP₂, $F_1(1, 41) = 7.39$, $p < .05$, and $F_2(1, 33) = 4.72$, $p < .1$, and at AP₃, $F_1(1, 41) = 4.09$, $p < .1$, and $F_2(1, 33) = 4.77$, $p < .1$, with no significant difference between priming effects at AP₂ and AP₃ ($F_1 < 1$ and $F_2 < 1$). Responses to long word targets, in contrast, showed no sign of facilitation and, in fact, tended to be slowed following test primes, though this interference was nonsignificant at AP₂ ($F_1 < 1$ and $F_2 < 1$) and only reached marginal significance by items at AP₃, $F_1(1, 41) = 1.89$, $p > .1$, and $F_2(1, 33) = 4.18$, $p < .1$. Comparisons of error rates showed that there were fewer lexical decision errors for short words following test primes than when following control primes at AP₂, $F_1(1, 41) = 9.11$, $p < .01$, and $F_2(1, 33) = 9.20$, $p < .01$. All other comparisons of error rates were nonsignificant (all $ps > .1$).

Combined analysis. RT data from Experiments 5A and 5B were z -score transformed as before, and test-control differences were entered into two-way ANOVAs with factors of target type (short or long word) and probe position. There was a highly significant effect of target type, $F_1(1, 186) = 15.55$, $p < .001$, and $F_2(1, 38) = 7.94$, $p < .01$, reflecting greater priming of short target words across all four probe positions. This finding suggests that embedded words were significantly more active than longer competitors throughout these non-garden-path sequences. The main

effect of probe position was nonsignificant by participants but marginal by items, $F_1 < 1$, and $F_2(3, 114) = 2.39$, $p < .1$, and there was no significant interaction between target type and probe position ($F_1 < 1$ and $F_2 < 1$). As in Experiment 2, this suggests that interpretations of non-garden-path stimuli (as measured by the magnitude of repetition priming) did not alter as participants heard more of the prime sentences.

Comparison of Experiments 2 and 5. Comparing the results obtained in this experiment with those for short word prime stimuli from Experiment 2 allowed us to investigate whether the time course of activation of short words and longer competitors was affected by the timing of mismatch between short word sequences and longer competitors. In the garden-path sequences used in Experiment 2, mismatch was delayed until the vowel of the following word, whereas in the non-garden-path sequences in Experiment 5, mismatch with longer competitors occurred immediately after the offset of the short word. In comparisons of priming effects for garden-path and non-garden-path stimuli, we used z -transformed difference scores as before.

We carried out three-way ANOVAs to compare the priming of different targets (short vs. long words) from different prime stimuli (short words followed by garden-path or non-garden-path continuations) over the four probe positions tested in Experiments 2 and 5. In analyses by items, each of these three independent measures was a repeated measures factor, whereas in analyses by participants, both prime type and probe position were between-groups comparisons.

These analyses showed a significant main effect of target type, $F_1(1, 400) = 15.40$, $p < .001$, and $F_2(1, 38) = 5.45$, $p < .05$, indicating greater overall priming for short targets than for long targets. Greater priming of short target words was unsurprising given that both sets of prime stimuli involved in this comparison

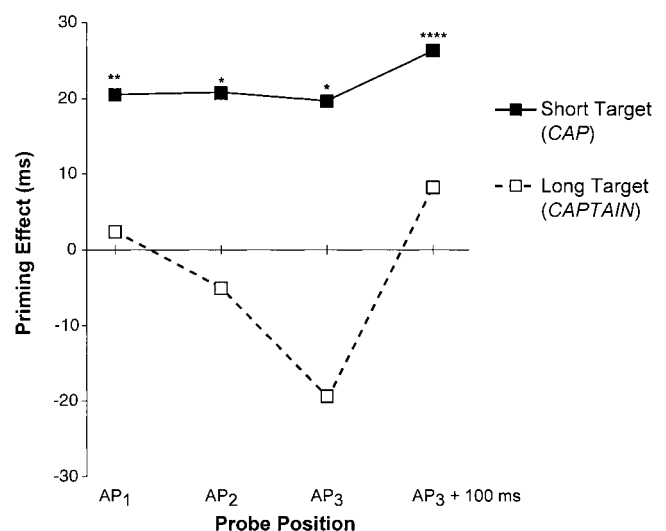


Figure 5. Experiments 5A and 5B: Magnitude and significance of priming (difference between response times following test and control primes) for short (CAP) and long (CAPTAIN) targets from non-garden-path sequences (cap looking). Experiment 5A: AP₁ and AP₄ (AP₃ + 100 ms). Experiment 5B: AP₂ and AP₃. AP = alignment point. Asterisks indicate significant priming in comparisons of responses following test and control primes. * $p < .10$. ** $p < .05$. *** $p < .01$. **** $p < .001$.

actually contained short words. A significant main effect of prime type (garden path vs. non-garden path) was also observed in these analyses, $F_1(1, 400) = 6.85, p < .01$, and $F_2(1, 38) = 11.96, p < .001$. The total magnitude of priming was greater for the lexical garden-path sequences than for the non-garden-path sequences. Furthermore, there was a significant interaction between the factors of prime and target type, $F_1(1, 400) = 2.85, p < .1$, and $F_2(1, 38) = 4.96, p < .05$. The interaction between prime type and target type indicates that both short and long targets were primed by garden-path sequences, whereas priming was only observed for short word targets for the non-garden-path sequences. There was also a marginally significant interaction between prime type and probe position, though this was nonsignificant by participants, $F_1(3, 400) = 1.30, p > .1$, and $F_2(3, 114) = 2.29, p < .1$. All other main effects and interactions were nonsignificant in this analysis ($F < 1$).

We carried out pairwise comparisons of priming effects to evaluate whether priming from garden-path and non-garden-path sequences differed for the two target types at each of the four probe positions. Hence, four comparisons were carried out for each target type (between four pairs of data points), with alpha levels corrected using the modified Bonferroni procedure (Keppel, 1982) such that each comparison reached statistical significance at a probability threshold of $p < .038$.

For long targets there was no difference in the magnitude of priming at AP₁, $F_1(1, 111) = 1.23, p > .1$, and $F_2(1, 38) = 2.03, p > .1$, or 100 ms after AP₃ ($F_1 < 1$ and $F_2 < 1$). However, at intermediate probe positions early in the continuation word, there were significant differences in the priming of long word targets from garden-path and non-garden-path stimuli. Significantly greater priming was observed for long word targets from garden-path primes at AP₃, $F_1(1, 102) = 7.77, p < .01$, and $F_2(1, 38) = 9.43, p < .01$, with a marginally significant difference at AP₂, $F_1(1, 94) = 3.02, p < .1$, and $F_2(1, 38) = 5.38, p < .05$. This pattern suggests that the following context of lexical garden-path stimuli increases the activation of long word competitors above the level of activation observed for non-garden-path sequences. For sequences such as *cap looking*, longer competitors do not appear to be activated after the offset of the embedded word. Thus, information after the offset of the embedded word acts to support or disconfirm longer lexical hypotheses. The implications of this finding are elaborated in the General Discussion section.

In contrast, however, the priming of short word targets in garden-path and non-garden-path stimuli did not differ at any of the probe positions tested. Pairwise comparisons contrasting garden-path and non-garden-path sequences were nonsignificant throughout (all $F_s < 1$), except for AP₃, $F_1(1, 102) = 1.75, p > .1$, and $F_2(1, 38) = 1.31, p > .1$. Although we are naturally cautious in interpreting null results in these comparisons, garden-path continuations did not appear to affect the magnitude of priming for short word targets, contrasting with the gating results for the same stimuli.

General Discussion

The results of these experiments provide detailed information about the time course of identification of onset-embedded words in connected speech. Embedded words have been crucial in motivating accounts of lexical segmentation that propose that recognition

of words in connected speech is often delayed until after their acoustic offset (see, for instance, Mattys, 1997; Norris, 1994).⁶ Many accounts of spoken word recognition (as well as the dictionary searches that have informed these accounts; see Luce, 1986; McQueen et al., 1995) assume that the lack of marked word boundaries in the speech stream produces ambiguity between short words and the long lexical items in which they are embedded. The experiments reported in this article were a direct test of this prediction, using stimuli designed to maximize ambiguity between short and long words. The embedded words used in Experiments 1, 2, and 3 were placed in nonbiasing sentential contexts and in phonological contexts that matched the longer competitors—the lexical garden paths referred to in the title of this article.

Word Boundary Ambiguity in Spoken Word Recognition

Experiments 1, 2, and 3 produced the novel result that even early in the processing of the critical words, our test stimuli were not as ambiguous as predicted by a phonemic analysis. Responses to the short and long word stimuli differed from the earliest gates tested in Experiment 1, long before the stimuli diverged phonemically. Results obtained in Experiments 2 and 3 provided further confirmation that the activation of short and long lexical items is biased by acoustic cues that differentiate short words from the initial syllables of longer words. In all four parts of Experiment 2, long targets were primed more strongly by long word primes than by short word primes, even when short and long word primes were phonemically identical (such as the syllable /kæp/ priming *CAPTAIN*). The converse finding for the priming of short target words was observed at two of the four probe positions tested in Experiment 2.

The results of Experiments 4 and 5 extend these findings by investigating the identification of short words in non-garden-path contexts. Results suggest that when the following context for a short word immediately mismatches with longer lexical items, little or no activation of longer competitors is observed. Thus, the partial activation of long target words for short word stimuli in Experiment 2 may be an artifact created by following contexts that temporarily match a longer word.

Differences in the priming of long target words by garden-path and non-garden-path contexts were observed at probe positions at which segments from the following context could be heard. In the gating experiment, significant differences were also observed at a probe position placed at the offset of the short word. The effect of mismatching information coarticulated from the following syllable is such that there is little ambiguity between embedded words and longer competitors, even at the offset of an embedded syllable.

These results further suggest that the degree of ambiguity cre-

⁶ Few accounts of spoken word identification explicitly propose a delay in identification; however, accounts of adult segmentation that use the phonotactics or metrical stress of following syllables as a cue to a word boundary are effective only for sequences in which multiple lexical candidates remain active after the offset of a word. These nonlexical cues to segmentation could therefore be used by adults only if identification is not achieved before the offset of a word or for sequences that contain novel or unfamiliar words.

ated by long words that contain onset embeddings has been overestimated by previous accounts of spoken word recognition. Additional cues are present in the speech stream that assist the perceptual system in distinguishing short words from the longer competitors in which they are embedded. As discussed previously, because many different acoustic cues could have been present in our stimuli, these experiments are inconclusive with respect to questions of the specific acoustic differences that are necessary or sufficient for discriminating short and long words. However, because significant correlations were observed between differences in duration, F_0 , and listener's ability to discriminate short and long word stimuli at the offset of the embedded syllable, our findings suggest that further investigation of these cues is merited.

Sequential Recognition and Lexical Competition

In the introductory section of this article, we distinguished two accounts of how lexical identification can contribute to segmentation. First, we described sequential recognition accounts, such as the original form of the Cohort model (Marslen-Wilson & Welsh, 1978), in which early identification allows a word boundary to be predicted after the offset of the current word. The inability of listeners to identify all words before their acoustic offset has been argued (Bard et al., 1988; Grosjean, 1985) to rule out sequential recognition accounts and motivate models of spoken word recognition, such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994), that incorporate lexical competition to divide the speech stream into words (see McQueen et al., 1995, for further elaboration).

The presence of acoustic cues that distinguish embedded words from the onset of longer competitors suggests that the mere presence of onset-embedded words in languages similar to English does not make delayed recognition obligatory in models of spoken word recognition. However, because our experiments also demonstrate effects of information appearing after the end of embedded syllables, post-offset influences on word recognition clearly need to be taken into account. For instance, in short word sequences, such as *cap tucked*, we saw activation of long lexical items after the offset of an embedded word. Only at the fourth probe position, around 200 ms after the offset of the embedded word, were short words primed without accompanying facilitation of long targets. Post-offset priming of long lexical items would not be predicted by accounts in which words are identified and alternatives ruled out before the offset of a word.

Comparisons between garden-path and non-garden-path sequences tested in Experiments 4 and 5 further illustrate the effect of information coming after the end of a word. Comparisons of Experiments 2 and 5 showed significant differences in the priming of longer competitors in garden-path and non-garden-path sequences at AP_2 and AP_3 . This finding indicates that information after the end of an embedded word can act to rule out longer competitors. These results demonstrate the effect of mismatching information in the speech stream.

Although these findings argue against accounts of lexical segmentation in which words are recognized and alternatives ruled out before word offset, we hesitate before concluding that these results support accounts of spoken word recognition that incorporate lexical-level competition. Differences between the recognition of

embedded words in garden-path and non-garden-path contexts indicate that mismatching information in short word sequences is used to rule out longer lexical candidates—long words are ruled out more rapidly if mismatching input appears earlier in the speech stream. Indeed, through coarticulation, it is possible that longer competitors can be ruled out before the offset of an embedded word—at least for stimuli without garden-path following contexts. These results are consistent with prior work showing that single feature deviations at either the beginning (Connine, Blasko, & Titone, 1993; Marslen-Wilson et al., 1996; Marslen-Wilson & Zwitserlood, 1989) or the end of a word (Gaskell & Marslen-Wilson, 1996; Marslen-Wilson et al., 1995) are sufficient to disrupt the lexical access process.

Our results extend these findings and demonstrate an effect of mismatch for stimuli in which the competitors being ruled out do not share the same word boundary as the target word (see Gaskell & Marslen-Wilson, 1996, for related work investigating mismatch created by phonological contexts that are nonviable for assimilation). Therefore, we propose that information in the speech signal that mismatches with activated lexical candidates not only plays a role in lexical selection at the single word level but also assists in deciding between candidates that span word boundaries, thereby assisting the segmentation of connected speech.

Models of spoken word recognition have typically postulated two different computational mechanisms by which mismatching input can rule out lexical candidates. In Shortlist and the Distributed Cohort model (Gaskell & Marslen-Wilson, 1997b, 1999), input that mismatches with currently activated items can directly reduce the activation of inappropriate lexical candidates. Conversely, in TRACE, no inhibitory connections exist between the speech input and lexical representations, and lexical selection operates solely through competition at the lexical level. Thus, in TRACE, mismatching input can only rule out a lexical item if an alternative word becomes more strongly activated. Although there is no reason to assume that these two mechanisms are mutually exclusive (indeed Shortlist contains both bottom-up inhibition and lexical-level competition), it appears models that include inhibitory connections between phonemes and lexical items that do not contain those sounds are to be favored: They can more readily simulate the results of experiments that show that mismatch that creates a nonword has an equally deleterious effect on lexical activation as mismatch that creates an alternative word (Frauenfelder, Segui, & Dijkstra, 1990; Marslen-Wilson & Gaskell, 1992; Marslen-Wilson & Warren, 1994).

The results of our experiments provide a further illustration of how mismatching information in the speech signal serves to rule out lexical candidates. For models such as TRACE that do not include these inhibitory connections, a decrease in the activation of long words can only be achieved by a corresponding increase in the activation of competing interpretations—most notably an increase in the activation of short embedded words. In the comparison of garden-path and non-garden-path stimuli in Experiments 2 and 5, no such change in priming of short word targets was observed as long words were ruled out. As we are cautious in drawing strong conclusions from this negative result, follow-up investigations to test this prediction of competition-based accounts are merited.

Modeling Spoken Word Identification

The results obtained in the current series of experiments suggest a speech perception system in which detailed acoustic information plays an important role in the lexical activation process. Instead of passing through initial stages that discard information by transforming the acoustic input into an abstract phonological or syllabic code, lexical access uses the full detail of the speech signal in identifying words (cf. Andruski, Blumstein, & Burton, 1994; Marslen-Wilson & Warren, 1994; Warren & Marslen-Wilson, 1987, 1988). We suggest that models such as TRACE and Shortlist, in which input representations are phonemically categorized, would be unable in their current form to simulate the experimental data reported in this article. These models would require alterations to incorporate the additional nonphonemic information that appears to bias lexical activation in choosing between short and long words. An alternative approach that we discuss proposes that there is a more direct mapping from the acoustic form of speech to lexical-semantic representations than there is in these phoneme-based accounts.

One model that exemplifies this direct-access approach to spoken word identification is the Distributed Cohort model (Gaskell & Marslen-Wilson, 1997b, 1999). In this account, speech perception is modeled as the process by which a continuous stream of speech is mapped onto lexical representations, without a mediating level of phonemically categorized representations. Implementations of this theory using recurrent neural networks have been shown to simulate the effect of subcategorical mismatch on the lexical access process (Gaskell & Marslen-Wilson, 1997b; Marslen-Wilson & Warren, 1994; however, see Norris, McQueen, & Cutler, 2000). They have also been able to simulate effects of lexical competition and bottom-up mismatch without directly implemented inhibitory connections between lexical units (Gaskell & Marslen-Wilson, 1999). However, the networks reported by Gaskell and Marslen-Wilson (1997b, 1999) were unable to track lexical activations across word boundaries and were consequently incapable of identifying short words embedded in longer items. More recent simulations are capable of reconciling the maximally efficient recognition of long words with the correct identification of short words embedded in longer items (Content & Sternon, 1994; Davis, Gaskell, & Marslen-Wilson, 1997; Davis, Marslen-Wilson, & Gaskell, 2000).

The crucial difference in the simulations described by Davis et al. (1997, 2000) is that these networks are trained not only to activate a representation of the current word in the sequence, but also to preserve activations representing previous words. Because the network must continue to activate words after their offset, short words will remain active long enough for their following context to rule out longer competitors. Simulations, in which these networks are provided with inputs that code for acoustic differences between syllables in short and long words, show a good match with the cross-modal priming data that was presented here (see Davis et al., 2000).

The experiments reported in this article provide detailed information regarding the time course of identification of onset-embedded words and longer competitors. Results suggest that the recognition system is sensitive to acoustic cues that differentiate syllables in short and long words. These acoustic cues greatly reduce the ambiguity created by short words embedded in longer

words. This pattern of results supports accounts of spoken word recognition that use the full detail available in the speech stream in activating lexical candidates. Computational models of spoken word recognition should therefore pay greater attention to the acoustic and informational properties of the speech input to simulate the processes involved in lexical segmentation and spoken word identification.

References

- Anderson, S., & Port, R. F. (1994). Evidence for syllable structure, stress and juncture from segmental durations. *Journal of Phonetics*, 22, 283–315.
- Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effects of subphonetic differences on lexical access. *Cognition*, 52, 163–187.
- Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Mahwah, NJ: Erlbaum.
- Baayen, R. H., Piepenbrock, R., & Guilikers, L. (1995). The CELEX Lexical Database (Version 2.5). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- 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 & Psychophysics*, 44, 395–408.
- Barry, W. J. (1981). Internal juncture and speech communication. In W. J. Barry & K. J. Kohler (Eds.), *Beiträge zur experimentellen und angewandten Phonetik* (pp. 231–288). Kiel, Germany: AIPUK.
- Brent, M. R. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences*, 3, 294–301.
- Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61, 93–125.
- Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus based approach to speech segmentation. *Cognitive Psychology*, 33, 111–153.
- Christiansen, M. H., & Allen, J. (1997). Coping with variation in speech segmentation. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of GALA 1997: Language acquisition. Knowledge representation and processing* (pp. 327–332). Edinburgh, Scotland: University of Edinburgh.
- Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221–268.
- Christie, W. M., Jr. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America*, 55, 819–821.
- Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 130–163). Hillsdale, NJ: Erlbaum.
- Connine, C., Blasko, D., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193–210.
- Content, A., & Sternon, P. (1994). Modelling retroactive context effects in spoken word recognition with a simple recurrent network. In A. Ram & K. Eiselt (Eds.), *Proceedings of the sixteenth annual conference of the Cognitive Science Society* (pp. 207–212). Hillsdale, NJ: Erlbaum.
- 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., & 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.
- Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, 128, 165–185.
- Davis, M. H., Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Recognising embedded words in connected speech: Context and competition. In J. A. Bullinaria, D. W. Glasspool, & G. Houghton (Eds.), *Proceedings of the fourth Neural Computation and Psychology workshop: Connectionist representations* (pp. 254–266). London: Springer-Verlag.
- Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2000). *Lexical segmentation and ambiguity: Investigating the recognition of onset-embedded words*. Paper presented at the Spoken Word Access Processes workshop, Nijmegen, Holland.
- Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, 30, 1–25.
- Forster, K. I., & Forster, J. C. (1995). DMASTR/VMASSTR [Computer software]. Retrieved from <http://www.u.arizona.edu/~kforster/dmastr>
- Forster, K. I., & Forster, J. C. (1999). DMDX (Version 1.0) [Computer software]. Retrieved from <http://www.u.arizona.edu/~kforster/dmdx/dmdxhp.htm>
- Frauenfelder, U. H., & Peeters, G. (1990). Lexical segmentation in TRACE: An exercise in simulation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 50–86). Cambridge, MA: MIT Press.
- Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 77–91.
- Gaskell, M. G. (1994). *Spoken word recognition: A combined computational and experimental approach*. Unpublished doctoral thesis, Birkbeck College, University of London, London, England.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144–158.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997a). Discriminating local and distributed models of competition in spoken word recognition. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference of the Cognitive Science Society* (pp. 247–252). Mahwah, NJ: Erlbaum.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997b). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12, 613–656.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition and blending in spoken word recognition. *Cognitive Science*, 23, 439–462.
- Gow, D. J., Jr., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 344–359.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 267–283.
- Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, 38, 299–310.
- Grosjean, F. (1996). Gating. *Language and Cognitive Processes*, 11, 597–604.
- Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25, 135–156.
- Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language*, 3, 367–382.
- Harris, Z. S. (1955). From phoneme to morpheme. *Language*, 31, 190–222.
- Isel, F., & Bacri, N. (1999). Spoken word recognition: The access to embedded words. *Brain and Language*, 68, 61–67.
- Keppel, G. (1982). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.
- Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208–1221.
- Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, 5(Suppl. 5), 5–54.
- 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 & Psychophysics*, 39, 155–158.
- Luce, P. A., & Cluff, M. S. (1998). Delayed commitment in spoken word recognition. *Perception & Psychophysics*, 60, 484–490.
- Luce, P. A., & Lyons, E. A. (1999). Processing lexically embedded spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 174–183.
- Luce, P. A., & Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Luce, P. A., Pisoni, D., & Goldinger, S. D. (1990). Similarity neighbourhoods 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.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, United Kingdom: Cambridge University Press.
- Marslen-Wilson, W. D. (1984). Function and processing in spoken word recognition: A tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 125–150). Hillsdale, NJ: Erlbaum.
- 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., & Gaskell, G. (1992). Match and mismatch in lexical access. *International Journal of Psychology*, 27, 61.
- Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1376–1392.
- Marslen-Wilson, W. D., Nix, A., & Gaskell, G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes*, 10, 285–308.
- Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101, 3–33.
- Marslen-Wilson, W. D., & 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.
- Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: On the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 576–585.
- Mattys, S. L. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin & Review*, 4, 310–329.
- Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics*, 62, 253–265.
- Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999).

- Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465–494.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McQueen, J. (1996). Word spotting. *Language and Cognitive Processes*, 11, 695–699.
- McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39, 21–46.
- McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, 10, 309–331.
- McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 621–638.
- Mertus, J. (1989). *BLISS user's manual*. Providence, RI: Brown University.
- Mertus, J. (1990). Brown Lab Interactive Speech System (Version 1.0) [Computer software]. Retrieved from <http://www.mertus.org>
- Monnell, S., & Hirsh, K. W. (1998). Competitor priming in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1495–1520.
- Moss, H. E., & Marslen-Wilson, W. (1993). Access to word meanings during spoken language comprehension: Effects of sentential semantic context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1254–1276.
- Moss, H. E., McCormick, S., & Tyler, L. K. (1997). The time-course of semantic activation in spoken word recognition. *Language and Cognitive Processes*, 12, 695–731.
- Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62, 715–719.
- 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.
- Norris, D. (1990). A dynamic-net model of human speech recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 87–104). Cambridge, MA: MIT Press.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.
- Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1209–1228.
- Norris, D., McQueen, J. M., & Cutler, A. (2000) Merging information in spoken word recognition: Feedback is never necessary. *Behavioural and Brain Sciences*, 23, 299–370.
- Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34, 191–243.
- Perruchet, P., & Vinter, A. (1998). PARSER: A model of word segmentation. *Journal of Memory and Language*, 39, 246–263.
- Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 785–794.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical language learning by 8 month olds. *Science*, 274, 1926–1928.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.
- 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.
- Swinney, D., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across modalities in the processing of individual words and sentences. *Memory & Cognition*, 7, 159–165.
- Tabossi, P. (1996). Cross-modal semantic priming. *Language and Cognitive Processes*, 11, 569–576.
- Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, 34, 440–467.
- Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). In W. B. Kleijn & K. K. Paliwal (Eds.), *Speech coding and synthesis* (pp. 495–518). New York: Elsevier.
- Toothaker, L. E. (1991). *Multiple comparisons for researchers*. Newbury Park, CA: Sage.
- Tyler, L. K. (1984). The structure of the word initial cohort: Evidence from gating. *Perception & Psychophysics*, 36, 417–427.
- Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics*, 34, 409–420.
- Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, 58, 434–445.
- 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.
- Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 710–720.
- Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word-spotting. *Memory & Cognition*, 24, 744–755.
- Warren, P., & Marslen-Wilson, W. D. (1987). Continuous uptake of acoustic cues in spoken word-recognition. *Perception & Psychophysics*, 41, 262–275.
- Warren, P., & Marslen-Wilson, W. D. (1988). Cues to lexical choice: Discriminating place and voice. *Perception & Psychophysics*, 43, 21–30.
- Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.
- Wolff, J. G. (1977). The discovery of segmentation in natural language. *British Journal of Psychology*, 68, 97–106.
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25–64.
- Zwitserslood, P. (1996). Form priming. *Language and Cognitive Processes*, 11, 589–596.
- Zwitserslood, P., & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, 10, 121–136.

(Appendix follows)

Appendix

Stimulus Sentences

Target words (short–long) appear in capitals, followed by stimulus sentences in the following order: (a) short word test stimuli (Experiments 1 and 2); (b) long word test stimuli (Experiments 1 and 2); (c) short word control prime stimuli (Experiment 2); (d) long word control prime stimuli (Experiment 2); (e) short word test prime with mismatching continuation (Experiments 3 and 4). Critical words are underlined.

ANT–ANTLER

- (a) It is because the ant lived under the rocks that it survived the explosion.
- (b) It is because the antler is fully grown that you can tell the deer is male.
- (c) It is because the horn was so loud that we all jumped.
- (d) It is because the trumpet was so loud that we all jumped.
- (e) It is because the ant found its way into the kitchen that we had to fumigate.

BAN–BANDAGE

- (a) Mike explained that the ban dates from the late 1930s.
- (b) Mike explained that the bandage was very tight in order to stop the bleeding.
- (c) Mike explained that the arch had been built by the Romans.
- (d) Mike explained that the cabbage always tasted horrible.
- (e) Mike explained that the ban solved the drinking problem.

BILL–BUILDING

- (a) It was agreed that the bill doesn't have to be paid immediately.
- (b) It was agreed that the building doesn't have to be pulled down immediately.
- (c) It was agreed that the name of the ship would be the Titanic.
- (d) It was agreed that the programme was hardly worth watching.
- (e) It was agreed that the bill for food should be paid immediately.

BOWL–BOULDER

- (a) We were lucky that the bowl didn't break when it hit the floor.
- (b) We were lucky that the boulder didn't crush us to death when it rolled down the hillside.
- (c) We were lucky that the rope didn't break with our combined weight.
- (d) We were lucky that the hammer was kept in the toolbox.
- (e) We were lucky that the bowl matched the one that we'd broken earlier.

BRAN–BRANDY

- (a) Susan claimed that the bran didn't taste nearly so bad.
- (b) Susan claimed that the brandy tasted much nicer.
- (c) Susan claimed that the chrome would never tarnish.
- (d) Susan claimed that the cupboard was much cheaper in the sale.
- (e) Susan claimed that the bran tasted much nicer.

CAN–CANTEEN

- (a) Opening the can takes a long time with a rusty penknife.
- (b) Opening the canteen was the cook's first job in the morning.
- (c) Opening the barn let the sheep out into the field.
- (d) Opening the hostel on a Sunday was a good idea.
- (e) Opening the can shouldn't take long with the right tool.

CAP–CAPTAIN

- (a) The soldier saluted the flag with his cap tucked under his arm.
- (b) The soldier saluted the flag with his captain looking on.
- (c) The soldier saluted the flag with his palm facing forwards.
- (d) The soldier saluted the flag with his rifle by his side.
- (e) The soldier saluted the flag with his cap looking slightly crumpled.

CHAP–CHAPLAIN

- (a) During the speech, the chap laughed at all the jokes.
- (b) During the speech, the chaplain started snoring really loudly.
- (c) During the speech, the hum died down.
- (d) During the speech, the platform started creaking alarmingly.
- (e) During the speech, the chap shut his eyes and went to sleep.

CREW–CRUSADE

- (a) It was unfortunate that the crew celebrated their victory so loudly.
- (b) It was unfortunate that the crusade was so violent.
- (c) It was unfortunate that the fog was so thick.
- (d) It was unfortunate that the garage was closed at weekends.
- (e) It was unfortunate that the crew veered into the bank at the start of the race.

CROW–CROQUET

- (a) After the lawn was mowed the crow could continue looking for food.
- (b) After the lawn was mowed the croquet match could begin.
- (c) After the lawn was mowed the weeds could be seen more clearly than ever.
- (d) After the lawn was mowed the picnic could take place.
- (e) After the lawn was mowed the crow gave up looking for worms.

CRY–CRISIS

- (a) Everyone was worried as the cry seemed to come from the attic.
- (b) Everyone was worried as the crisis was getting worse by the minute.
- (c) Everyone was worried as the exam was much harder than expected.
- (d) Everyone was worried as the engine had started making loud noises.
- (e) Everyone was worried as the cry didn't sound like it came from the TV.

DEN–DENTIST

- (a) At the end of a hard day, the den tends to be the place I choose to relax.
- (b) At the end of a hard day, the dentist needed somewhere to relax.
- (c) At the end of a hard day, the chores are the last thing I want to do.
- (d) At the end of a hard day, the washing up is the last thing I want to do.
- (e) At the end of a hard day, the den should be an ideal place to relax.

DOCK–DOCTOR

- (a) On Saturdays the dock teemed with people.
- (b) On Saturdays the doctor was always very busy.
- (c) On Saturdays the ducks are usually very well fed.
- (d) On Saturdays the circus is fully booked.
- (e) On Saturdays the dock should be fairly quiet.

DOLL–DOLPHIN

- (a) The children thought the doll felt softer than usual.
- (b) The children thought the dolphin was beautiful.
- (c) The children thought the clown was very funny.
- (d) The children thought the museum was very boring.
- (e) The children thought the doll could be fun to play with.

FAN–FANCY

- (a) Everyone agreed that the fan suited Catherine's new outfit.
- (b) Everyone agreed that the fancy clothes suited Catherine.
- (c) Everyone agreed that the bait should be suitable for catching rats.
- (d) Everyone agreed that the rations were inadequate for adults.
- (e) Everyone agreed that the fan should be left on during the afternoon.

GIN–GINGER

- (a) A splash of gin just about makes the drink perfect.
- (b) A splash of ginger makes whiskey taste really good.
- (c) A splash of soup ruined my outfit.
- (d) A splash of curry ruined my outfit.
- (e) A splash of gin tastes really good with ice and lemon.

GREY–GRAVY

- (a) Some time later, the grey van was all that people talked about.
- (b) Some time later, the gravy was all that people talked about.
- (c) Some time later, the feast began to get livelier.
- (d) Some time later, the dagger was found.
- (e) Some time later, the grey car was all that people talked about.

HAM–HAMSTER

- (a) During the summer it is best if the ham stays in the fridge.
- (b) During the summer it is best if the hamster stays in the shade.
- (c) During the summer it is best if the shrubs are watered regularly.
- (d) During the summer it is best if the moped is kept in the garage.
- (e) During the summer it is best if the ham never gets left out of the fridge.

HELL–HELMET

- (a) The soldiers thought that hell might be more comfortable than their barracks.
- (b) The soldiers thought that helmets would save their lives.
- (c) The soldiers thought that tents wouldn't stay dry if it rained.
- (d) The soldiers thought that aeroplanes were the best way to travel.
- (e) The soldiers thought that hell tormented the souls of their enemies.

JUNK–JUNCTION

- (a) It was obvious that the junk should be moved somewhere else.
- (b) It was obvious that the junction was dangerous to drive around.
- (c) It was obvious that the gems weren't worth very much money.
- (d) It was obvious that the cider was much stronger than usual.
- (e) It was obvious that the junk made the house look less tidy.

KID–KIDNEY

- (a) We were concerned when the kid knocked over the priceless vase.
- (b) We were concerned when the kidney infection hadn't got any better.
- (c) We were concerned when the flight was delayed by a couple of hours.
- (d) We were concerned when the bouquet of flowers didn't arrive.
- (e) We were concerned when the kid laughed at violent movies.

LAWN–LAUNDRY

- (a) On sunny days, the lawn dried out, leaving large brown patches.
- (b) On sunny days, the laundry was hung out in the garden to dry.
- (c) On sunny days, the bay was crowded with holidaymakers.
- (d) On sunny days, the canyon was filled with haze.
- (e) On sunny days, the lawn tends to be covered with people sunbathing.

NAP–NAPKIN

- (a) Taking a nap can help you to stay up later.
- (b) Taking a napkin from the restaurant was a good idea.
- (c) Taking a dip in the sea is very nice during the summer.
- (d) Taking a hostage allowed the robbers to make their escape.
- (e) Taking a nap tends to help me stay up later.

PAIN–PAINTING

- (a) John replied that the pain tempted him to abort the climb.
- (b) John replied that the painting was very colourful.
- (c) John replied that the songs were quite good.
- (d) John replied that the record was quite good.
- (e) John replied that the pain wouldn't stop him from climbing.

PAN–PANTRY

- (a) Although he was an experienced cook, the pan transformed Bruce's cooking.
- (b) Although he was an experienced cook, the pantry contained ingredients Bruce had never seen before.
- (c) Although he was an experienced cook, the sauce was a real challenge to get right.
- (d) Although he was an experienced cook, the onions still made him cry when he chopped them.
- (e) Although he was an experienced cook, the pan saved Bruce a lot of trouble.

PEN–PENSION

- (a) We all noticed that the pen shook when the young man signed the form.
- (b) We all noticed that the pension payments were worth less and less each month.
- (c) We all noticed that the skirt didn't match Anne's blouse.
- (d) We all noticed that the trousers didn't match Peter's jacket.
- (e) We all noticed that the pen changed Phillip's handwriting for the better.

PIG–PIGMENT

- (a) Because of its odd appearance, the pig made everyone gasp with astonishment.
- (b) Because of its odd appearance, the pigment was rejected by Dulux.
- (c) Because of its odd appearance, the tie attracted attention.
- (d) Because of its odd appearance, the bicycle was never stolen.
- (e) Because of its odd appearance, the pig never got sold at market.

PILL–PILGRIM

- (a) They hoped that the pill granted them immunity from the disease.
- (b) They hoped that the pilgrim would save them.
- (c) They hoped that the hint would be understood.
- (d) They hoped that the basement would not get flooded by the storm.
- (e) They hoped that the pill didn't have any unpleasant side effects.

POLE-POULTRY

- (a) As we climbed over the farm gate, the pole tripped us up.
- (b) As we climbed over the farm gate, the poultry ran away from us.
- (c) As we climbed over the farm gate, the heel on my shoe came loose.
- (d) As we climbed over the farm gate, the orchard could be seen.
- (e) As we climbed over the farm gate, the pole didn't support our weight.

SHELL-SHELTER

- (a) Although badly battered, the shell tempted the collector.
- (b) Although badly battered, the shelter was warm and dry.
- (c) Although badly battered, the yacht was still watertight.
- (d) Although badly battered, the vessel was still watertight.
- (e) Although badly battered, the shell might still be valuable.

SPY-SPIDER

- (a) We had to be careful that the spy didn't overhear our conversations.
- (b) We had to be careful that the spider didn't crawl into our sleeping bags.
- (c) We had to be careful that the jeans were washed inside out.
- (d) We had to be careful that the ferry was on time.
- (e) We had to be careful that the spy listened to the fake recording.

STAY-STATION

- (a) They thought that the stay became boring after a while.
- (b) They thought that the stable would cost more than the house to heat.
- (c) They thought that the kiln was hot enough to fire the pots.
- (d) They thought that the pistol belonged to the criminal.
- (e) They thought that the stay ceased being interesting after the first week.

TRACK-TRACTOR

- (a) When it reached the house, the track turned north towards the forest.
- (b) When it reached the house, the tractor came to a halt.
- (c) When it reached the house, the cat was offered a saucer of milk.
- (d) When it reached the house, the parcel remained unopened for several days.
- (e) When it reached the house, the track got more difficult to follow.

TRAY-TRAITOR

- (a) After a while, the tray tempted him too much and he started to eat.
- (b) After a while, the traitor became careless and he was caught.
- (c) After a while, the flag was raised to the top of the flagpole.
- (d) After a while, the kettle came to the boil.
- (e) After a while, the tray should have been returned to the kitchen.

TREE-TREATY

- (a) For the last fifty years there has been a tree towering above this house.
- (b) For the last fifty years there has been a treaty between England and Germany.

- (c) For the last fifty years there has been a race to see who could climb the hill fastest.
- (d) For the last fifty years there has been a butcher's in the high street.
- (e) For the last fifty years there has been a tree standing on this spot.

TRY-TRIFLE

- (a) We were disappointed that the try failed to win the match.
- (b) We were disappointed that the trifle hadn't been touched.
- (c) We were disappointed that the queen didn't come to visit the school.
- (d) We were disappointed that the princess didn't come to visit the school.
- (e) We were disappointed that the try very nearly lost us the match.

WALL-WALNUT

- (a) A severe storm left the wall nearest the house badly damaged.
- (b) A severe storm left the walnut tree badly damaged.
- (c) A severe storm left the town with a large bill for the clear-up operation.
- (d) A severe storm left the locals with a large bill for the clear-up operation.
- (e) A severe storm left the wall teetering on the brink of collapse.

WELL-WELCOME

- (a) In the village, the well can't cope with this summer's drought.
- (b) In the village, the welcome given to tourists is very friendly.
- (c) In the village, the fumes from the factory are unbearable.
- (d) In the village, the parson is very friendly.
- (e) In the village, the well might not cope with this summer's drought.

WIN-WINTER

- (a) After a bad start to the season, the win turned the teams' fortunes around.
- (b) After a bad start to the season, the winter became much milder than usual.
- (c) After a bad start to the season, the drought was eased by the arrival of the monsoon.
- (d) After a bad start to the season, the public stopped attending the matches.
- (e) After a bad start to the season, the win helped our team to avoid relegation.

WIT-WITNESS

- (a) Everyone thought Tom's wit nearly deserved a prize.
- (b) Everyone thought Tom's witness was the least convincing.
- (c) Everyone thought Tom's socks were a horrible colour.
- (d) Everyone thought Tom's jacket made him look very smart.
- (e) Everyone thought Tom's wit made him an ideal companion for the trip.

Received June 10, 1999
 Revision received March 26, 2001
 Accepted June 20, 2001 ■