Discriminating Local and Distributed Models of Competition in Spoken Word Recognition

M. Gareth Gaskell and **William D. Marslen-Wilson**
Centre for Speech and Language,
Psychology Department, Birkbeck College,
Malet Street, London WC1E 7HX, England
{g.gaskell,w.marslen-wilson}@psyc.bbk.ac.uk

Abstract

Local and distributed theories of representation make different predictions regarding the simultaneous activation of multiple lexical entries during speech perception. We report three experiments that use the cross-modal priming technique with fragments of spoken words to explore competition effects in the activation of multiple lexical representations. The experiments suggest that lexical activation is inversely related to the number of words being activated. This competition effect is stronger at the semantic than the phonological level of representation, supporting a model of speech perception in which sensory information is mapped directly onto distributed representations of both the form and the meanings of words.

Introduction

Current models of cognitive functioning often describe representations with respect to a local/distributed distinction. A pure localist system uses a unique processing element for each item to be represented. At the other extreme, a fully distributed system uses each element to represent every item. In many situations, this distinction is more a matter of perspective than a hard theoretical division. Indeed if the items in an distributed system can be represented by unbounded linearly independent vectors, the system is isomorphic to a localist system (Smolensky, 1986). However, when this assumption does not hold, some identifiable differences do occur.

Gaskell (1996) examined the combinatorial properties of localist and distributed systems, with the specific intention of modeling simultaneous activation of lexical representations during spoken word recognition. A number of cross-modal priming studies have demonstrated that lexical knowledge about more than one word is transiently activated during the perception of a spoken word. For example, Marslen-Wilson (1987) showed that the presentation of a spoken word fragment such as *capt...* facilitates the recognition of words associated with more than one possible continuation (e.g., *ship* related to *captain*, *guard* related to *captive*).

These, and other similar data, suggest that a model of lexical access in speech perception must allow multiple word representations to be activated simultaneously. This finding is easily captured in a localist model, which simply activates the nodes representing each of the words that match the fragment so far. For a distributed model it is not so simple, because overlapping patterns of activation will create interference. Standard connectionist networks, when faced with this problem, will produce a "blend" pattern, equivalent to the arithmetic mean of the vectors representing the patterns to be activated. Blending may be best appreciated using a spatial analogy. Distributed patterns can be represented by points in a multidimensional state space, with each dimension of the space representing a single element in the distributed pattern. The blend of a set of patterns is simply their midpoint in this space. Blending is a rather noisy solution to the problem of simultaneous activation: as the number of patterns to be activated increases, the similarity of the blend to each pattern decreases and the blend may indeed turn out to be more similar to some other pattern. A behavioral analogy to this problem would be creating a blend of the distributed representations of *captain* and *captive* in response to the fragment *capt...* and accessing the meaning of, say, *banana*, which happens to be close to the blend in the state space.

Fortunately, this possibility recedes as the number of dimensions in the representational space increases. Nonetheless, the basic finding of Gaskell (1996) was that distributed representations can only be activated simultaneously at a cost, and that for some critical number of items multiple activation would become noisy or even break down altogether.

An extension to this basic finding indicates that the ability to represent multiple items simultaneously in a distributed system depends on the organization and distribution of the items within the system. If the system is organized so that all items that are likely to be activated together have similar representations, then their blend is also likely to be similar to the constituent items and thus represent those items more successfully. In the case of spoken words, this is true if lexical representations are organized in terms of their phonology. If, in some sub-area of lexical space, representational dimensions correspond either explicitly or implicitly to phonological features, then lexical blends of words like *captain* and *captive* will be similar to those words and relatively dissimilar to phonologically unrelated words. On the other hand, because words with similar onsets tend to be unrelated in meaning, this advantage would not hold for a distributed lexicon in which items are semantically organized (see Figure 1). Gaskell and Marslen-Wilson (1995, in press) proposed a model of lexical access in which a featural representation of speech is mapped directly onto a distributed representation incorporating both lexical form and meaning. This model predicts that the ability to activate words in parallel will vary for different areas of lexical space. This variation should be observable by manipulating the experimental task to focus on either semantic or phonological organization of the mental lexicon.

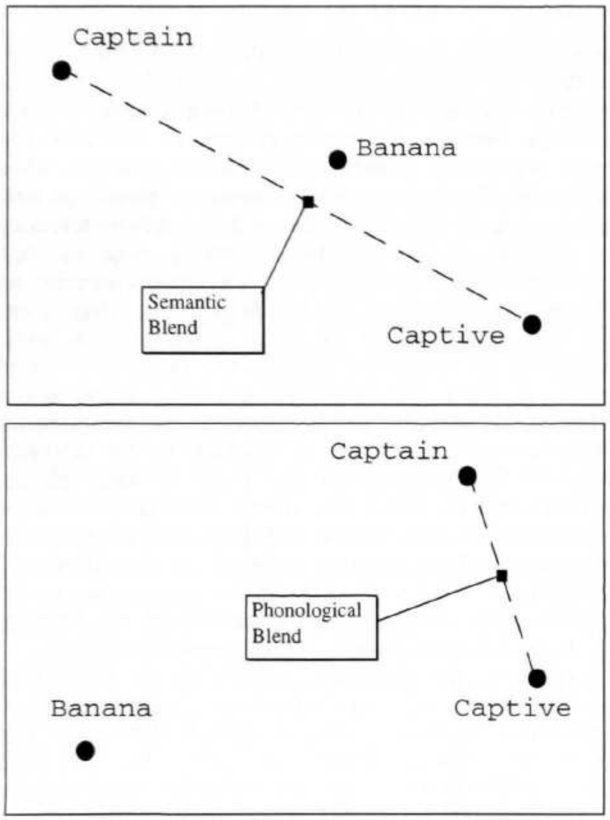

Figure 1. Blending of onset-matched word representations for spaces organized in terms of meaning (upper box) and phonology (lower box). The filled circles represent lexical items and the filled squares represent the blend in each type of space. The representations of unrelated words, such as banana, are more likely to be similar to the blend in semantic space than in phonological space.

Returning briefly to localist models of lexical representation, in which a word is represented by the activation of a single node, simultaneous representation can occur with no cost in terms of reduced activation. Regardless of the number of lexical matches to a fragment of speech, the nodes representing each lexical item can be independently fully activated (e.g., Marslen-Wilson & Welsh, 1978). However, although localist models *can* have this property, they do not have to. For example, the TRACE model of speech perception (McClelland & Elman, 1986) builds in inhibitory links between localist word nodes in order to capture competition between word candidates. Localist models are less constrained in these terms, implying that it is easier to find evidence supporting localist over distributed models than the converse. Nonetheless the prediction of distributed models—that the organization of words within the lexicon affects the ability to activate word representations in parallel—is not shared by TRACE or any other localist model of lexical access in speech.

Previous Research

A number of studies (e.g., Marslen-Wilson, 1987; Zwitserlood, 1989) have addressed multiple activation in speech perception using associative priming, for which the target is an associate of the prime (e.g., elbow-grease), chosen from normative data (e.g., Moss & Older, 1996). Recent computational and experimental work has suggested that this type of relationship may be influenced predominantly by knowledge about the cooccurrence of word forms and is thus less useful for assessing access to stored lexical knowledge (Moss, Hare, Day & Tyler, 1994; Plaut, 1995). More recently, similar questions have been addressed using a non-associative semantic relationship between prime and target (e.g., Moss, McCormick & Tyler, in press; Zwitserlood & Schriefers, 1994). These have generally shown weak priming effects for fragments of words cut off before uniqueness point, but have not directly compared the levels of facilitation found for fragments of speech matching different numbers of words.

Zwitserlood and Schriefers (1994) demonstrated the importance of prime duration in the interpretation of priming strength in terms of lexical activation. Their experiment compared the priming effects of a short fragment of a word to either a longer fragment of the same word or simply the short fragment followed by an inter-stimulus interval equivalent to the difference in duration between the short and the long fragment (roughly 100 ms on average). The short fragments did not facilitate recognition of semantically related targets, but the other two prime types produced equal amounts facilitation. This suggests that there is a delay between information becoming available and the observation of its effects at a lexical level. It also demonstrates the problems involved in trying to find effects of competition on activation by using different length fragments of the same word as stimuli in a priming experiment, because of the confound of processing time. Instead, different sets of words, matched for fragment duration and varying in competitor environment, must be used.

Experimental Issues

Our current research examines the effects of competition on the degree of activation of lexical candidates using the priming technique. Two sets of words were created, which varied in the number of possible continuations at certain points. For one group, the words were uniquely identifiable before the final consonant or consonants (e.g., /gɑmənt/, *garment*, where no other word matches the fragment /gɑmə/). We shall label these words the early uniqueness point (UP) group. For the second group (late UP), at least

Table 1: Example primes for each condition, with phonemic transcription. The numbers of words that onset-match each related prime phonemically (averaged across all items in that category) are given in brackets.

	Related prime			Unrelated control		
	CVC(C)	CVC(C)V	Complete	CVC(C)	CVC(C)V	Complete
Early UP	garm (23)	garme (1)	garment(1)	chis	chise	chisel
	/gɑm/	/gɑmʌ/	/gɑmʌnt/	/tʃɪz/	/tʃɪzʌ/	/tʃɪzʌl/
Late UP	capt (39)	captai (3.5)	captain (1)	mount	mountai	mountain
	/kæpt/	/kæptɪ/	/kæptɪn/	/mawnt/	/mawntɪ/	/mawntɪn/

two possible continuations were possible at this point (e.g., /kæptɪn/, *captain*, where the fragment /kæptɪ/ is consistent with both *captain* and *captive*). According to distributed models of lexical access, as well as localist models incorporating competition, the complete activation of the words in the late UP set should only occur once the final consonants of the words are heard. Before this point, more than one candidate should be partially activated. On the other hand, for the early UP group complete activation is possible earlier on in the perception of the word. Localist models that do not introduce direct competition between word nodes (e.g., Cohort; Marslen-Wilson & Welsh, 1978) predict no difference in activation between the two groups before the final vowels, since they allow multiple word candidates to be fully activated.

Distributed models further predict variation in activation depending on the organization of the section of lexical space under examination. Specifically, a distributed phonological representation provides a better basis for parallel activation than a distributed semantic representation, because the lexical items that must be coactivated during speech perception will be similar to each other phonologically, but not semantically. To explore this issue, we varied the relationship between prime and target between experiments. For Experiment 1, the target was the orthographic form of the prime and for Experiment 2, the target was a close category coordinate of the prime. The issue of what levels of representation these tasks tap into is not uncontroversial, but it seems likely that cross-modal repetition priming will be more influenced by lexical phonological form than semantic priming.

Materials and Gating Pre-test

All experiments employed the same materials and design (see Table 1). There were two within-item variables that manipulated the form of the spoken prime: prime length (the length of the fragment of speech used as prime) and prime-target relatedness (either related or unrelated). The three levels of prime length involved the bisyllabic spoken prime word being presented either with the final vowel and consonant(s) spliced out, or with the final consonant(s) spliced out, or as a complete word. Each related prime token was matched with a similarly spliced unrelated control prime. A third independent variable involved the manipulation of competitor environment between items (early vs late UP words) as discussed above. The prime words were concrete nouns, matched on phonemic length, temporal duration and frequency of usage. In all priming experiments, the prime word was presented spoken, in isolation, and the visual target was presented at the offset of the speech. The task of the participant was to make a timed *yes/no* lexical decision to the target.

The initial estimate of competitor environment was made by phonemic dictionary search using the MRC machine readable dictionary. Discounting morphological variants, the prime words were all uniquely identifiable when complete. At the second level of prime length (i.e., with the final consonant(s) of the word removed), the related primes in the early UP condition were still uniquely identifiable, but the late UP related primes were now consistent with at least 2 different word continuations. At the first level of prime length (i.e., with the final vowel and consonant(s) removed) there were generally more words consistent with the fragments for both groups of fragments, with a range across groups of 1 to 289. The mean number of possible continuations for each related prime condition are given in Table 1.

Although the phonemic dictionary searches were useful for selecting stimuli, they tend to overestimate the number of matches between a speech fragment and lexical items. This is because firstly they do not specify subtle information in the speech signal (e.g., coarticulatory information) that might be used to discriminate between lexical candidates, and secondly they include rare words that may not be known to the majority of language users. In order to obtain a secondary measure of competitor environment, a reduced form of the gating task (Grosjean, 1980) was used in which participants were presented the related spoken primes in order of prime length (e.g., /kæpt/, followed by /kæptɪ/, followed by /kæptɪn/). Their task was to write down what they though the complete word was at each point and to rate the confidence of their decision.

Mean percentage identification rates and confidence values (out of 10) are displayed in Figure 2. The pattern of results agrees quite closely with the dictionary estimates of competitor environment, although the percentage of correct responses early in the word is greater than the dictionary searches would suggest. For the complete words, both sets of words are identified perfectly. However, for both the spliced conditions, there is a strong effect of competitor environment, with early UP words correctly identified more often than the late UP words. The ratings data display this competition effect more weakly, with a stronger overall effect of gate position.

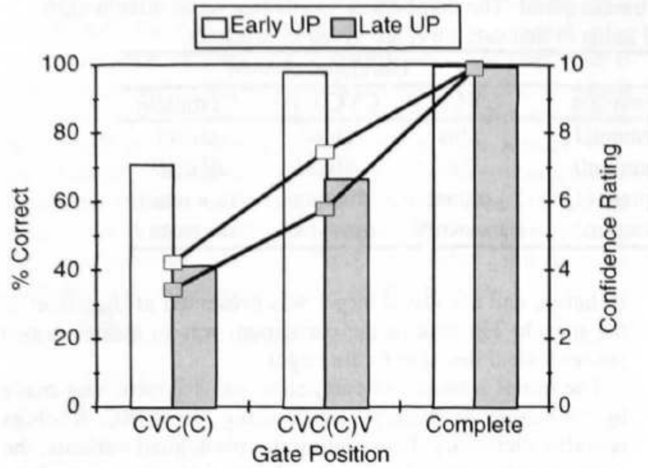

Figure 2. Results of the gating pre-test. The bars correspond to the mean identification percentages (left hand scale) and the lines correspond to the mean confidence ratings (right hand scale).

Experiment 1: Repetition Priming

Experiment 1 uses repetition priming to examine the effects of competition on the activation of words during speech perception. The model of Gaskell and Marslen-Wilson (1995) predicts that there should be more evidence of coactivation of multiple lexical entries using a task that involves phonological as well as or instead of semantic dimensions of lexical space. We assume that repetition priming is suitable in this respect.

The materials described above were used as primes in a cross-modal priming experiment. On each trial, an auditory prime was presented followed immediately by a visual target word, which for the test items was the complete orthographic form of the related prime in capital letters (e.g., CAPTAIN or GARMENT). Participants were instructed to make a two alternative forced choice lexical decision on the target word. Interspersed with the test items were a large number of filler trials, which ensured that the ratio of word to nonword targets was equal and reduced the prime-target relatedness proportion to 29%. Filler items with nonword targets related in form to the prime (e.g., *secret*-SECREN) prevented form links between primes and targets being used as cues to word targets.

The results of Experiment 1 are summarized in Figure 3 in terms of the amount of priming found in each condition. This is the difference between the mean response times to the related and unrelated conditions for each combination of prime length and competitor environment. A positive value suggests that the related prime has facilitated the lexical decision to the target word, and further suggests that the lexical representation of that word has been activated. Modified Bonferroni planned comparisons showed that the facilitatory effects of the primes were significant in all conditions. This implies that even quite short fragments of speech, which match the onsets of many words, will facilitate the recognition of those words. Previous experiments have shown that this effect disappears if the lexical item and the spoken prime mismatch to even a slight extent. For example, under similar conditions, *heap* or *heak* will not facilitate the recognition of *heat* (Marslen-Wilson, Nix & Gaskell, 1995). The facilitation cannot be explained simply on the basis of a general phonological overlap between different lexical items. Instead, the results are consistent with a model of lexical access in speech perception in which multiple lexical items can be activated in parallel based on their lexical form.

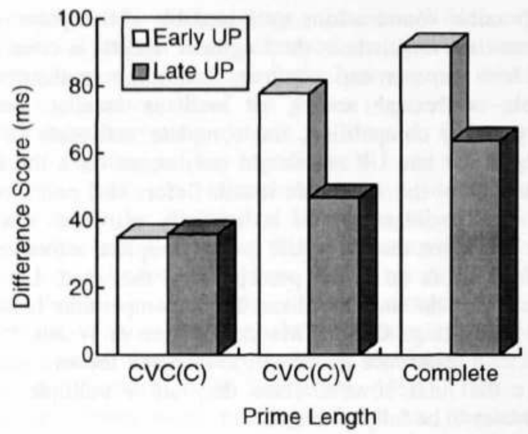

Figure 3. Repetition priming results

Despite the overall facilitation found in Experiment 1, there is also evidence suggesting that the degree of facilitation depends on the number of lexical items matching the spoken prime. This is evident firstly in the general increase in facilitation as more of a prime word is heard and the number of matching lexical items drops ($F_1[2, 96] = 9.5$, $p < .001$; $F_2[2, 138] = 9.2$, $p < .001$), although an alternative explanation of this finding could be that the increasing duration of the primes accounts for the increased facilitation (cf. Zwitserlood & Schriefers, 1994). More crucially, the difference between the amounts of facilitation for the early and late UP items was significant for the CVC(C)V stimuli, which were matched in terms of duration and number of phonemes ($F_1[1, 48] = 7.7$, $p < .01$; $F_2[1, 69] = 4.0$, $p < .05$). This is an advantage of tokens that match a unique lexical item over tokens that match at least two items. If we interpret strength of facilitation as a measure of degree of activation of the target word, then it seems that multiple activation has a cost. Complete activation of a lexical form representation relies on the deselection of all competitors, whereas the simultaneous activation of more than one lexical item implies a partial activation.

Experiment 2: Semantic Priming

Experiment 2 applies the design and prime materials of Experiment 1 to a different set of target words, for which the target was semantically, but not associatively, related to prime. The targets were chosen from the WordNet database (Miller, 1990) and were near synonyms or close category coordinates of the primes (e.g., *captain-commander*). Our intention in Experiment 2a was to examine how parallel activation of multiple lexical items affects access to the meanings of those words. Apart from the replacement of form-related nonword filler trials, which were no longer required, the breakdown of experimental lists was identical to that of Experiment 1. The results, in the form of priming values are displayed in Figure 4.

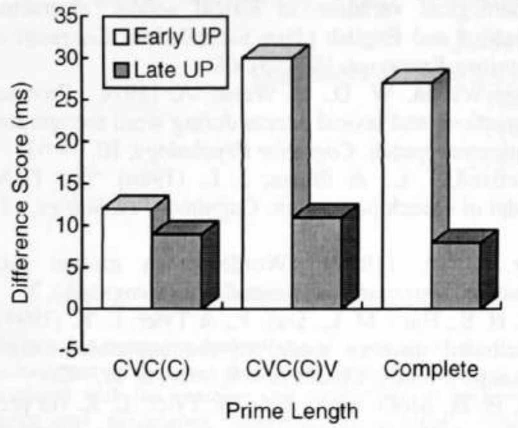

Figure 4. Semantic priming results

Compared to Experiment 1, priming in all conditions was greatly reduced. In fact the priming levels only reached significance for the complete and final consonant excised conditions of the early UP words, when the prime word is uniquely identifiable. For the late UP items, not even the complete words produced significant priming of the related targets. Thus there is no evidence from Experiment 2a for parallel activation of lexical semantic representations. Like Experiment 1, there is evidence of competition effects: there was significantly more facilitation for the early than the late UP words overall ($F_1[1, 48] = 4.7$, $p < .05$; $F_2[1, 69] = 3.6$, $p = 0.06$). However, the lack of priming even at the offset of the late UP words is surprising. It is possible that the strength of competition involved for these words delays access to meaning until after the offset of the word. Such a lag between speech information entering the perceptual system and the resolution of competition was observed by Zwitserlood and Schriefers (1994) and is predicted by models such as TRACE (McClelland & Elman, 1986), which require repeated cycles of processing for mismatching auditory information to take effect at the word level.

Experiment 2b, which replicated Experiment 2a with a delay of 250 ms between the offset of the prime and the onset of the target, confirmed this hypothesis (see Figure 5). All conditions showed similar levels of priming to Experiment 2a apart from the complete late UP condition, which in Experiment 2b showed significant priming. There is a detectable delay between auditory information entering the perceptual system and the consequences of the new information showing up at the semantic level.

Discussion

In all three priming experiments we find some effect of competitor environment that modulates the strength of facilitation of target words. Spoken word fragments that match a single lexical item reliably facilitate the recognition of the visual form of that word and also generally facilitate recognition of non-associated category coordinates of that word (with the exception of late UP words immediately followed by a semantically related word). These results indicate that a unique match between speech and a lexical form allows full activation of its representation and access to stored knowledge about the word. This finding is unsurprising and does not contradict any model of lexical access.

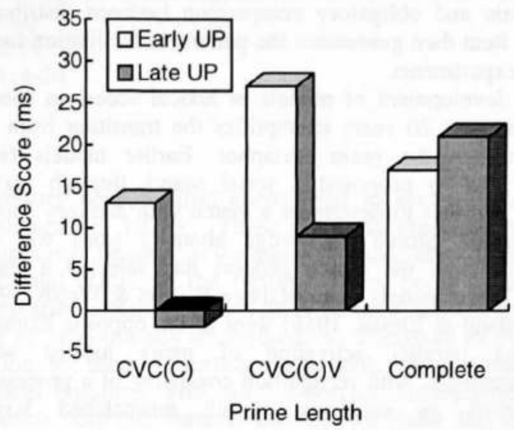

Figure 5. Delayed semantic priming results

More importantly, when there is still ambiguity about the identity of a spoken fragment we find weaker facilitation and in some cases no facilitation. The degree of facilitation (and thus the degree of activation of the matching lexical entries) seems to be inversely related to the number of word matches. This contradicts localist models, such as the original Cohort model (Marslen-Wilson & Welsh, 1978), that do not incorporate direct inhibition between lexical candidates. The competition effect is explained either by incorporating lateral inhibition in localist models (e.g., TRACE) or by "natural competition" (Smolensky, 1986) in distributed models (e.g., Gaskell & Marslen-Wilson, 1995).

Finally, comparing the results across experiments we find greater evidence for parallel activation in a task exploiting the forms of words than in one involving the meanings of words. Facilitation by non-unique word fragments, although partial, was reasonably strong in the repetition priming experiment, but became weaker or disappeared altogether when the targets were semantically related to the prime.[1] This result was predicted by the distributed model of Gaskell & Marslen-Wilson (1995, in press; see Gaskell, 1996, Figure 6). A blend of multiple distributed representations is more likely to be informative if those representations are similar to each other in the state space spanned by the distributed vectors. This is true when all or part of the distributed representations encode similarity of

[1] In our experiments, there was no significant semantic priming before uniqueness point. However, it is possible that a more powerful experimental design would pick up the weak effects (of the order of 10 ms) in the anticipated direction. This null result does not invalidate previous demonstrations of semantic priming using ambiguous word fragments (e.g., Zwitserlood & Schriefers, 1995).

phonological form and the selected words share a common onset.

We should note at this point that a variant of, say, TRACE that retained the localist word nodes and added a distributed semantic layer could quite probably accommodate these findings. Nonetheless the most parsimonious account of these data requires simply that acoustic input is mapped directly onto a single distributed space encompassing (at least) lexical semantics and lexical phonology. The resulting automatic and obligatory competition between distributed lexical item then guarantees the pattern of facilitation found in our experiments.

The development of models of lexical access in speech over the past 20 years exemplifies the transition from the computer to the brain metaphor. Earlier models (e.g., Forster, 1976) proposed a serial search through lexical entries in order to determine a match with auditory speech information. Stored knowledge about a word was not available until the search process had selected a single match. Later models (e.g., Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986) went to the opposite extreme, stressing parallel activation of many localist word representations, with recognition consisting of a process of elimination as incoming speech mismatched lexical candidates.

Our proposal, that speech is mapped directly onto a distributed representation of word forms and meanings, is in fact a compromise between these two positions. Parallel activation of multiple distributed representations is permitted in this type of model, with the proviso that complete activation of a full representation can only occur for a single lexical item. Lexical access proceeds as a gradual emergence of signal from noise, as lexical candidates are rejected and the blend of distributed patterns is adjusted to eliminate the mismatching representations.

Acknowledgements

This research was supported by a UK MRC programme grant awarded to William Marslen-Wilson and Lorraine Tyler. We thank members of the Birkbeck Centre for Speech and Language for useful suggestions.

References

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

Gaskell, M. G. (1996). Parallel activation of distributed concepts: who put the P in the PDP? In G. W. Cottrell (Eds.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Gaskell, M. G., & Marslen-Wilson, W. (1995). Modeling the perception of spoken words. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 19-24). Mahwah, NJ: Erlbaum.

Gaskell, M. G., & Marslen-Wilson, W. D. (in press). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes.*

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

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

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

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

Miller, G. A. (1990). "WordNet: an on-line lexical database". *International Journal of Lexicography*, 3.

Moss, H. E., Hare, M. L., Day, P., & Tyler, L. K. (1994). A distributed memory model of the associated boost in semantic priming. *Connection Science*, 6, 413-426.

Moss, H. E., McCormick, S. F., & Tyler, L. K. (in press). The time course of activation of semantic information during spoken word recognition: function precedes form. *Language and Cognitive Processes.*

Moss, H. E., & Older, L. (1996). *Birkbeck word association norms*. Hove: Erlbaum.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 37-42). Mahwah, NJ: Erlbaum.

Smolensky, P. (1986). Neural and conceptual interpretation of PDP models. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models*. Cambridge, MA: MIT Press/Bradford Books.

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

Zwitserlood, P., & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, 10(2), 121-136.