

7. Effects of following context in recognising embedded words

In previous chapters it has been shown that acoustic cues to word length are used in the identification of words embedded at the onset of longer words. In none of the repetition priming experiments reported in Chapter 5 was significant priming of onset-embedded words (such as *cap*) observed from stimuli containing longer words that contained these embeddings (*captain*) - even where the prime word was cut off at the offset of a syllable matching an embedded word. However, these experiments also showed that during the identification of sequences containing an embedded word with a garden-path continuation (such as the test sequence *cap tucked*), longer competitors (e.g. *captain*) remain active until after the offset of the embedded word. Significant priming of longer words was observed at probe positions up to the vowel of the following syllable.

In the recurrent network simulations reported in Chapter 6, the lack of ambiguity between short and long words was modelled by incorporating an input cue analogous to the duration difference between syllables in short and long words. The network was able to use this cue even where syllables of short and long words are presented with the same, ambiguous duration. This indicates that the network is able to process duration adaptively, in order to disambiguate onset-embedded words from longer competitors. The model is therefore able to account for priming results suggesting that short and long words are not as ambiguous as previously predicted. The network also simulated the identification of embedded word sequences that contain a lexical garden-path. Where continuations for short word stimuli matched the second syllable of a longer word, activations for the longer lexical item were boosted during identification. Consequently, the garden-path sequences that were used in Experiments 1 and 2 were responsible for the increased activation of longer words like *captain* in the recurrent network simulations.

For this reason, it is expected that short word stimuli would only cause ambiguity for sequences that create a lexical garden-path. Input sequences such as *cap lick* (shown in Figure 6.5a and b) that mismatch with all longer words do not activate longer competitors after the offset of the embedded word. It is therefore a crucial test of the validity of the

model to investigate whether listeners also show a different activation profile for sequences containing onset-embedded words in non-garden-path following contexts. The experiments reported in this chapter therefore investigated the identification of short embedded words in contexts that rule out longer interpretations immediately after the offset of the embedded word. The results of these experiments can then be compared with the activation profile predicted by the network.

7.1. Experimental materials

The goal of the experiments reported in this chapter was to investigate the time course of recognition of onset-embedded words under conditions where less disruption from longer competitors was predicted. Sequences containing an embedded word were used with a continuation that diverges from all other lexical items immediately after the offset of the embedded word (rather than the delayed mismatch that was used in Experiments 1 and 2). As before, both gating and repetition priming will be used to probe the activation of short and long target words at different points in the speech stream. Since the critical comparisons here involve short word stimuli, it is not necessary to re-run the long word materials a second time. Furthermore since short words with garden-path contexts have already been tested in Experiments 1 and 2 it should not be necessary to re-test these stimuli. The experiments reported here therefore used only a single set of test sentences containing short, embedded words, with continuations that immediately mismatch with all longer competitors.

The items used in this experiment were derived from the same set of onset-embedded monosyllables used in Experiments 1 and 2, placed in the same context sentences as before. Words following the monosyllabic test word were changed so that the onset segments of the following word mismatched with all likely longer competitors. For an example pair of test words *cap* and *captain*, other lexical items that start with the syllable [kæp] include: *caption*, *capsule*, *captive* and *capture*. The onset of the following word was therefore chosen to mismatch with all the longer lexical items in the CELEX database (Baayen, Popenbrook, & Guilikers, 1995). Since a continuation that begins with the phoneme /l/ mismatches with all these continuations, the sequence *cap looking* was used in the prime sentence for this stimulus item.

Results obtained by McQueen (1998) showed that listeners detected embedded words more easily where they are followed by a phonotactically illegal sequence. Continuations were therefore chosen such that the segments either side of the word boundary are found word-internally in other English words. These phonotactically legal sequences would not therefore provide a pre-lexical cue to a word boundary. For the example word *cap*, other words in the CELEX database with the segments /æp/ at the end of a syllable include: *clapboard*, *haphazard*, *napkin*, *chaplain*, *entrapment*. By analogy with the word *chaplain*, continuations starting with the segment /l/, would therefore not provide a phonotactic cue to a word boundary. Hence the test sentence used for the monosyllable *cap* was “*The soldier saluted the flag with his cap looking slightly crumpled*”. A similar process was repeated for each of the onset-embedded words used in Experiments 1 and 2. The complete set of 40 sentences is shown in Appendix A.

These materials were recorded and digitised using the same methods employed previously. Three alignment points, equivalent to those used in Experiments 1 and 2 were marked for these stimuli (AP_1 at the offset of the first syllable of the test word, AP_2 after the onset segments of the following word and AP_3 in the vowel of the second syllable). The durations of each of these sections were compared to those of the short-word stimuli used in Experiments 1 and 2 (as shown in Table 4.1 in Chapter 4). The target monosyllable (e.g. *cap*) averaged 303ms in duration, which does not significantly differ from the mean duration (291ms) of the equivalent syllable in Experiments 1 and 2 ($t(39)=1.32$, $p>0.1$). The onset segments of the second syllable (between AP_1 and AP_2) were 73ms in duration (compared to 78ms in Experiment 1 and 2, $t(39)=0.42$, $p>0.1$), and the third alignment point was placed an average of 43ms into the vowel of the second syllable (42ms previously, $t(39)=0.66$, $p>0.1$).

To allow comparison between the results of the current experiment and those reported previously, both gating and cross-modal repetition priming experiments were carried out using these stimuli. As in the previous experiments the goal of these investigations was to assess the ongoing competition between short and long word interpretations of these stimuli. With this in mind, methods were used to measure the activation of both the target word and a longer competitor. Since these experiments can be compared to those reported in Chapters 4 and 5, only the new set of short word stimuli are tested here.

7.2. Experiment 3 – Gating

The gating task was used initially to investigate whether the identification of embedded monosyllables follows a different pattern from that reported in Experiment 1 for sentences in which mismatch with longer competitors occurs in the onset of the following word.

Participants

Eleven subjects from the Birkbeck Speech and Language subject pool took part in the experiment. All were paid £5/hour for their participation. None had taken part in any of the previous experiments.

Design, materials and procedure

These were as described in Experiment 1 in Chapter 4, except that instead of stimuli containing either short or long target words, all the test stimuli contained embedded words with following contexts that immediately mismatched with longer lexical items. Gates were set up as previously at the three alignment points described in Figure 4.1 and Table 4.1 with two additional gates 50 and 100ms before AP_1 , and other gates 50, 100, 200, 300 and 400ms after AP_3 . Since only a single set of test sentences were investigated, these were presented in a single test version. As in Experiment 1, test stimuli were played out in successive fragments, preceded by 4 practice items and interspersed with 16 filler items. As previously the experiment was divided into 5 blocks – one of practice items and four blocks of test and filler items. The experiment took approximately two hours to complete, including short breaks following each block of test items.

7.2.1. Results and discussion

Results were analysed in terms of the proportion of responses at each gate that matched either the target word or a longer competitor. Three items produced a disproportionate number of errors. For *bran* and *ban* over 50% of participants failed to identify these words correctly by the final gate – responses for these items were therefore discarded. The third discarded item (*win*) produced a lower error rate (36%) in the current experiment but was discarded to aid comparisons with Experiment 1 (in which over 60% of participants failed to identify this target word). The proportion of responses matching either short or long target words at different gates is shown in Figure 7.1. Also included in this graph is

data from the short word stimuli in Experiment 1 where stimuli included lexical garden-paths after the offset of the short word.

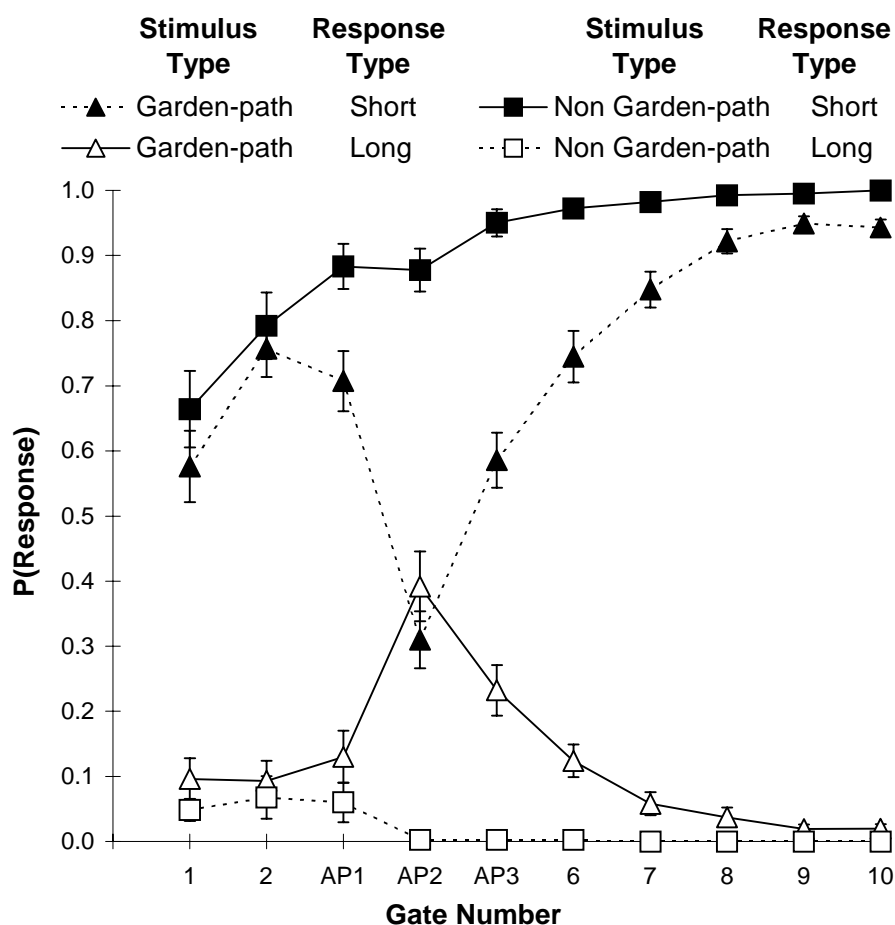


Figure 7.1: Results of Experiment 3 – Gating. Proportion of responses matching short (CAP) and long words (CAPTAIN) for stimuli containing lexical garden-paths (*cap tucked*) and without garden-paths (*cap looking*). Error bars = 1 standard error. Lexical garden-path data from Experiment 1 (Figure 4.2)

As can be seen in Figure 7.1, participants in this experiment no longer produced reduced numbers of short word responses at gates where the following word can be heard. This change is especially apparent at AP_2 where information about the onset of the following word becomes available. However, there is also a discrepancy at AP_1 between short word responses to stimuli with delayed or immediate mismatch ($t_1(31)=4.57$, $p<.001$; $t_2(37)=4.35$, $p<.001$). This indicates that for non garden-path stimuli additional information supporting the identification of onset-embedded words is available at AP_1 . This result suggests that the earliest alignment point used in these experiments does include some information from the following context. Given the co-articulation of

segments in connected speech it will be difficult to cut-off the end of a word in such a way as to exclude any influence from the onset of the following word.

The effect of garden-path sequences on the identification of embedded words in Experiment 1 can also be seen in the proportion of correct responses made at later gates. The results from the current experiment indicate that (apart from the excluded items) participants correctly identified all the target words at the final gate, whereas previously only 94% made at gate 10 correctly identified the embedded word in garden-path contexts. Significantly fewer correct responses ($t_1(31)=4.51$, $p<.001$; $t_2(35)=4.64$, $p<.001$) were produced for stimuli with continuations that match longer lexical items – even where mismatch is only delayed over the onset segments of the following syllable.

Another way of analysing this data is by calculating isolation points. These measure the gate at which participants produce the correct response without subsequently altering their response (Grosjean, 1996). In the analysis of Experiment 1, the isolation point of the short word stimuli was bi-modally distributed with some participants isolating the embedded word before AP_1 and some after AP_3 . In the current experiment a dip in the number of correct short-word responses is no longer observed at AP_2 . For this reason, measures of isolation point now provide an appropriate summary of the identification of the short word stimuli in non garden-path contexts. The mean isolation point for these stimuli in Experiment 3 (without *ban*, *bran* and *win*) was 268ms; which is shorter than the average duration of the target words (303ms). A paired t-test showed that these isolation points were significantly before the offset of the target word ($t(36)=3.29$, $p<.01$).

The results of this analysis suggests that these stimuli could be correctly identified at their offset - unlike the lexical garden-path stimuli that were used in Experiment 1. It is possible that other gating experiments that demonstrated delayed (i.e. post-offset) identification of words in connected speech (Bard, Shillcock, & Altmann, 1988; Grosjean, 1985) may have inadvertently contained lexical-garden paths equivalent to those that were deliberately constructed in the short word stimuli for Experiments 1 and 2. Alternatively it may be that the stimuli used in the current experiments provide a stronger contextual constraint than those used previously, producing earlier isolation points in the gating task (Tyler & Wessels, 1983).

7.3. Experiment 4 - Cross-modal priming

In the concluding section of Chapter 4, the issue of response biases in gating was discussed. These may limit the accuracy of the gating task as a means of gauging the lexical activation of words in connected speech. Consequently, further studies reported in Chapter 5 used cross-modal repetition-priming to provide an on-line measure of listeners' interpretations of stimuli containing onset-embedded words. The same method will now be applied to short word stimuli with non garden-path continuations. As in Experiment 2, repetition priming will be used to measure the lexical activation of embedded words and longer words that contain the embedded word as their initial syllable. In this way the extent of ambiguities between short and long words can be established for these non-garden-path stimuli, extending the results of Experiment 2.

As in Experiment 3, this study will only use a single set of test stimuli instead of the two sets used in Experiment 2. In order to assess the activation of targets and competitors at a single probe position, four experimental conditions are required (two prime types - test and control - and two target types - short and long words). To test all four probe positions separately as in Experiment 2 would therefore require four experiments each with four versions. To reduce the number of experimental versions (and hence number of subjects) required, control prime conditions were combined so that two probe positions for the test stimuli were tested together with a single control prime. This reduced the size of the experiment to 12 conditions which were tested in two separate six version experiments.

Participants

Over the two repetition-priming studies carried out in Experiment 4, 114 participants from the Birkbeck Speech and Language subject panel were tested (56 on Experiment 4a and 58 on Experiment 4b). A shortage of previously untested subjects meant that approximately 20 of the subjects had taken part in Experiment 2(a). However, none had been tested on these experiments within the previous 12 months and none had been tested on the stimuli used in the current experiment (i.e. none had taken part in Experiment 3). All were native English speakers without any hearing or language impairment and were paid for their participation.

Design

The 40 stimulus sentences containing short prime words with non-garden-path continuations were paired with the same set of control prime sentences in which the test word is replaced with an unrelated word matched in frequency to either the short or long target (see Appendix A). Since there was only one set of test sentences, two different probe positions were examined in an experiment with three prime types (two test primes and one control prime). To cover the four probe positions assessing the activation of both short and long words, two experiments with six versions were designed. Experiment 4a measured priming at AP_1 and AP_4 (100ms after AP_3) and Experiment 4b probed at AP_2 and AP_3 . Control primes in each experiment were played up to a point equivalent to the earlier of the two probe positions - AP_1 in Experiment 4a and AP_2 in Experiment 4b.

In all other respects both experiments were identical to each other. Each contained a total of 122 filler sentences, 27 of which were paired with a non-word phonologically related to the word at the cut-off point of the sentence, as well as 54 sentences followed by unrelated non-words and 41 followed by unrelated word targets. This produced experimental versions in which 50% of targets were words and where 33% of targets followed a phonologically similar prime. When the 20 practice items and 10 dummy items (starting each experimental block) are included, the proportion of trials in which a word target was preceded by a related test prime was just over 14%.

Procedure

The procedure used in Experiment 4 was identical to that used in Experiment 2 except for the different test stimuli used. These were presented up to AP_1 or AP_4 in Experiment 4a or up to AP_2 and AP_3 in Experiment 4b. Details of each test trial are as described for Experiment 2 with visual targets being presented for 200ms at the point where the speech is cut off. Once again, the experiment was divided into four sessions, an initial block of practice items, followed by two blocks of experimental items and finishing with a recognition memory test on some of the filler sentences used in the experiment.

7.3.1. Results

Response time data from these two experiments was analysed following the exclusion of slow or error prone participants (mean RT over 750ms or error rates of over 12.5% on the

test words). These criteria led to the exclusion of 8 participants from Experiment 4(a) and 11 participants in Experiment 4(b). As in Experiment 2, the target word BRAN elicited a large number of errors and it, along with its matched pair BRANDY was excluded from further analysis. Also excluded were a number of outlying responses slower than 1200ms (4 data-points from Experiment 4a and 1 data-point from Experiment 4b). Response times and error rates for each prime and target type following these exclusions are shown in Table 7.1 below.

Experiment 4a					Experiment 4b				
Prime Type <i>Probe</i>	Short Target (CAP)		Long Target (CAPTAIN)		Prime Type <i>Probe</i>	Short Target (CAP)		Long Target (CAPTAIN)	
	RT (ms)	error %	RT (ms)	error %		RT (ms)	error %	RT (ms)	error %
Test <i>AP₁</i>	488	2.7	558	5.5	Test <i>AP₂</i>	499	1.0	556	5.7
Test <i>AP₄</i>	480	1.6	549	7.6	Test <i>AP₃</i>	500	3.5	569	5.9
Control	508	4.5	560	7.5	Control	520	5.7	551	6.2

Table 7.1: Results of Experiment 4a and 4b. Mean lexical decision times and error rates by prime and target type.

Statistical analysis of this experiment used pairwise comparisons of responses following test and control primes to evaluate the magnitude and significance of priming effects. Analyses comparing the priming effects found in different experiments will use differences between normalised control and test prime RTs as the dependent measure. Analysis of response times and errors by prime and target type for each experiment are reported in Appendix B.

Experiment 4a (AP₁ and AP₄)

Responses to short word targets were significantly facilitated by test primes at AP₁ ($\underline{F}_1[1,42]=5.87$, $p<.05$; $\underline{F}_2[1,33]=5.00$, $p<.05$) and at AP₄ ($\underline{F}_1[1,42]=14.73$, $p<.001$; $\underline{F}_2[1,33]=13.62$, $p<.001$). No significant priming effects were observed for long targets (all $\underline{F}_1<1$ and $\underline{F}_2<1$). Differences in error rates following test and control primes failed to show any significant differences (all $p>.1$ – except differences in error rates to short words following test primes at AP₄ which was significant by participants $\underline{F}_1[1,42]=4.98$, $p<.05$; $\underline{F}_2[1,33]=2.35$, $p>.1$). The magnitude and significance of the priming effects observed for each prime and target type are shown in Figure 7.2.

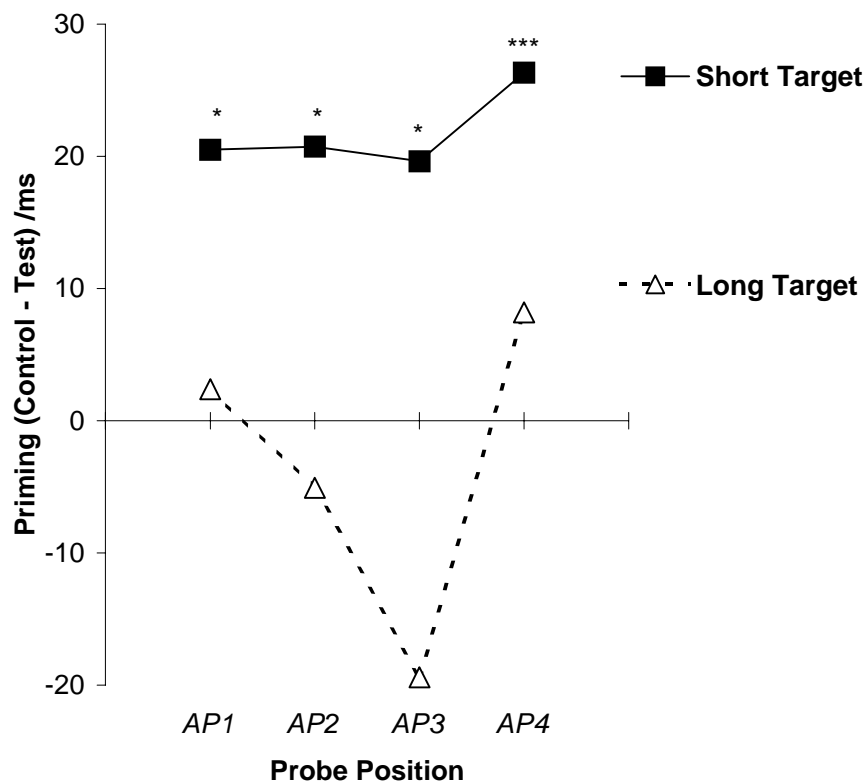


Figure 7.2: Magnitude and statistical significance of priming in Experiment 4. Primes contain short words with non-lexical continuations (*cap looking*) with short and long word targets (CAP/CAPTAIN) presented at different probe positions (AP₁, AP₂, AP₃, AP₄). *** $p<.001$; * $p<.05$

Experiment 4b (AP₂ and AP₃)

Pairwise comparisons of response times following test and control primes illustrated in Figure 7.2 show significant priming for short word targets at AP₂ ($\underline{F}_1[1,41]=7.39$, $p<.01$; $\underline{F}_2[1,33]=4.72$, $p<.05$) and at AP₃ ($\underline{F}_1[1,41]=4.09$, $p<.05$; $\underline{F}_2[1,33]=4.77$, $p<.05$). Responses to long word targets in contrast, tended to be slowed following test primes.

This interference effect was not significant at AP_2 ($F_1 < 1$, $F_2 < 1$) though it approached significance at AP_3 ($F_1[1,41]=1.89$, $p > .1$; $F_2[1,33]=4.18$, $p < .05$). Comparisons of error rates showed that there were fewer lexical decision errors for short words following test primes than following control primes at AP_2 ($F_1[1,41]=9.11$, $p < .01$; $F_2[1,33]=9.20$, $p < .01$). All other comparisons of error rates were non-significant (all $p > .1$).

Pairwise comparisons of errors made following test and control primes indicate that participants made fewer lexical decision errors for short words following test primes than following control primes. This effect was significant following primes presented up to AP_2 ($F_1[1,41]=9.11$, $p < 0.01$; $F_2[1,33]=9.20$, $p < 0.01$) though not at AP_3 ($F_1[1,41]=2.05$, $p > 0.1$; $F_2[1,33]=2.41$, $p > 0.1$). There was no significant difference in error rates to long word targets following test primes at either probe position compared to controls (all $F_1 < 1$ and $F_2 < 1$).

Combined analysis (Experiment 4a and 4b)

Combined analysis of priming effects in Experiment 4a and 4b were carried out using data that had been normalised by dividing each response time by the mean for that participant and multiplying by the overall mean response time in both experiments. Test-control difference scores for this normalised data were entered into two-way ANOVAs using the factors target type (short or long words) and probe position. There was a highly significant effect of target type ($F_1[1,186]=17.70$, $p < .001$; $F_2[1,38]=7.68$, $p < .01$) reflecting greater priming of short target words across all four probe positions (see Figure 7.2). There was no main effect of probe position ($F_1[3,186]=1.32$, $p > .1$; $F_2[3,114]=2.41$, $p < .1$) nor any significant interaction between target type and probe position ($F_1 < 1$; $F_2 < 1$). As in the analysis across different probe positions for Experiment 2, the lack of a significant effect of probe position suggests that interpretations of the stimuli (as indicated by the magnitude of priming) did not significantly change as participants heard more of the prime sentences.

7.4. Comparison of Experiment 2 and Experiment 4

Analyses of the two sets of repetition priming experiments were carried out to compare the magnitude of repetition priming for short word stimuli with garden-path and non-garden path following contexts. These ANOVAs used normalised difference scores as the

dependent measure, to minimise differences produced by variation between the groups of subjects tested in each part of the two experiments. Analyses with participants as the random variable included the prime type factor (garden-path vs. non garden-path stimuli) as a between participants comparison (since these are the results of separate experiments) while analyses by items were carried out with both prime and target type as repeated-measures factors. The magnitude of priming for short and long target words from embedded words with garden-path continuations (*cap tucked*) and stimuli without lexical garden-paths (*cap looking*) is shown in Figure 7.3.

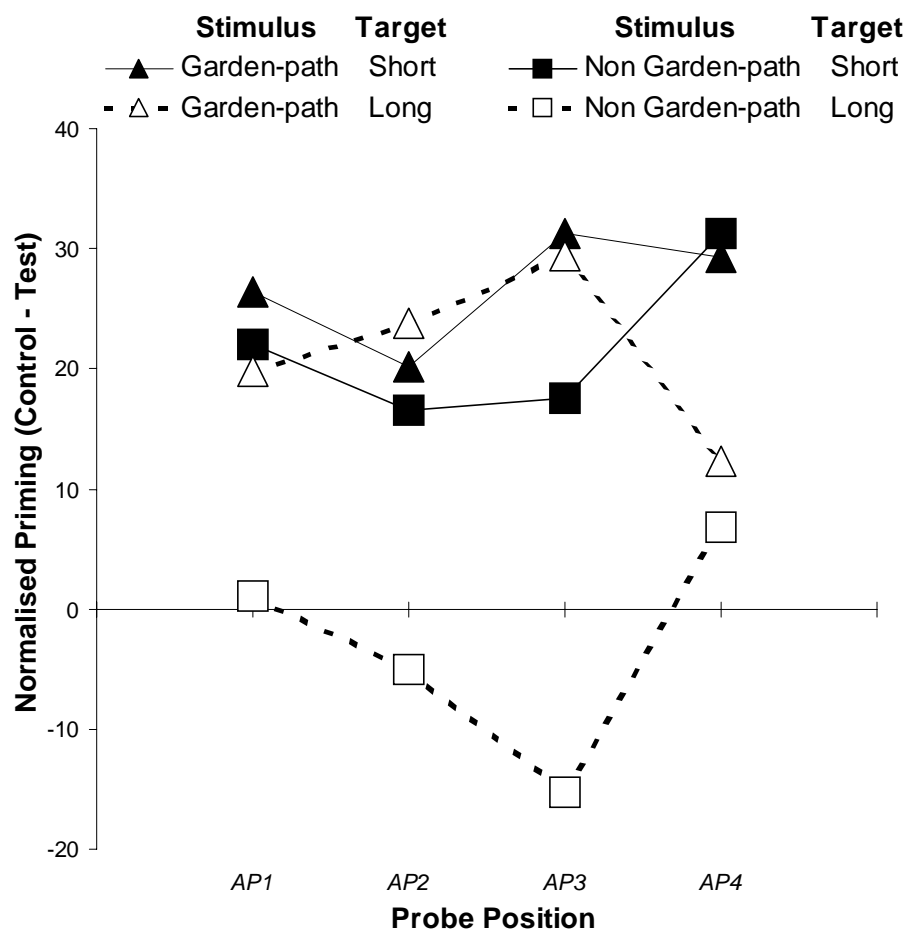


Figure 7.3: Combined results of Experiments 2 and 4. Normalised priming of short (CAP) and long (CAPTAIN) targets for garden-path stimuli (*cap tucked*) and non garden-path stimuli (*cap looking*) at different probe positions.

Analyses showed a significant main effect of target type ($F_1[1,400]=15.09$, $p<.001$; $F_2[1,38]=4.78$, $p<.05$) indicating greater overall priming for short target words than for long targets. This greater priming of short words is unsurprising given that both sets of prime stimuli involved in this comparison contained short words. A significant main

effect of prime type (lexical garden-path vs. non garden-path) was also observed in these analyses ($\underline{F}_1[1,400]=9.40, p<.01$; $\underline{F}_2[1, 38]=11.21, p<.01$). The total magnitude of priming was greater for the lexical garden-path sequences than the non garden-path sequences.

An interesting effect here is the interaction between prime type and target type - though this is only marginally significant in the analysis by participants ($\underline{F}_1[1,400]=2.86, p<.1$; $\underline{F}_2[1,38]=4.61, p<.05$). As Figure 7.3 shows, the lexical garden-path stimuli used in Experiment 2 produce significant priming of both short and long words at most probe positions, while the non garden-path stimuli in Experiment 4 only facilitate short word targets.

There was no main effect of probe position in this analysis, nor any interaction between probe position and prime or target type. However, pairwise comparisons at individual probe positions suggest that differences in the activation of long lexical items for stimuli with and without lexical garden-paths are observed at specific probe positions. At AP_1 and AP_4 there are no significant differences between the priming of long words in Experiments 2 and 4 (at AP_1 , $\underline{F}_1[1,111]=1.67, p>.1$; $\underline{F}_2[1,38]=2.75, p>.1$. At AP_4 , $\underline{F}_1<1$; $\underline{F}_2<1$). However, differences in the priming of long targets do emerge at AP_2 and AP_3 . Long word targets are primed more strongly in Experiment 2 than in Experiment 4 at AP_2 ($\underline{F}_1[1,70]=3.37, p<.1$; $\underline{F}_2[1,38]=5.05, p<.05$) and at AP_3 ($\underline{F}_1[1,102]=6.97, p<.01$; $\underline{F}_2[1,38]=6.93, p<.05$)

Thus for stimuli that mismatch with the longer word immediately after the offset of the embedded word (*cap looking*), reduced priming is observed for long targets earlier in the speech stream than for lexical garden-path primes (*cap tucked*). These effects of the presence or absence of garden-path continuations suggest that bottom-up mismatch acts to support or disconfirm alternative lexical hypotheses – even where the competing interpretations do not share word boundaries.

In contrast to the results for long targets, there were no significant differences in the priming of short words in garden-path and non garden-path stimuli at any probe position (all $p>.1$). Caution is required in interpreting null results in between-experiment comparisons, however the presence or absence of garden-path continuations appears to have no significant effect on the priming of short word stimuli – contrasting with the gating results presented previously. The implications of these results for models of spoken word recognition will be discussed in the concluding section of this chapter.

7.5. Comparing experimental data and recurrent network simulations

The recurrent network model developed in this thesis suggests that mismatch between continuations of embedded words and longer lexical items plays an important role in the identification of onset-embedded words. Simulation 1, reported in Chapter 3, did not incorporate any other input cue to distinguish syllables of short and long words. Consequently, lexical units for onset-embedded words were only fully activated when longer competitors were ruled out by mismatching following contexts. Thus, mismatching input provided the only cue that allowed onset-embedded words to be identified. The networks described in Chapter 3 therefore predict marked differences in the activation of short word units depending on whether the following context of an embedded word forms a lexical garden-path with a longer word or not (for an example, compare the activation profile for *cap* in Figure 3.4b and Figure 3.5b).

Since the simulations reported in Chapter 6, however, included an input cue analogous to the acoustic difference between syllables in short and long words. These networks no longer rely solely on post-offset mismatch to recognise embedded words. Therefore, it is unclear the extent to which networks in Simulation 4 still predict differences in activation for short word stimuli depending on their following context. Further tests of these networks were therefore carried out to compare the activation of short and long word units in response to stimuli containing short words followed by garden-path and non garden-path contexts. Results from these simulations can then be compared to that obtained from the combined priming data from Experiments 2 and 4.

Data for the garden-path sequences will be identical to that used in simulating the results of Experiment 2. The non-garden-path stimuli were newly generated sequences in which the initial segment following the offset of the embedded word mismatched with the longer lexical item that contained the embedded word. Since there was only one longer word containing each of the embedded monosyllables in the networks vocabulary these sequences were easier to generate than the equivalent experimental stimuli. All the stimuli included the embedded word as the second word in a sequence so that, although presented with an ambiguous duration, cues to the length of the target word could be detected by the network. The word before the test item was either monosyllabic or bisyllabic and matched

neither the embedded word nor its longer competitor. The test set contained 32 combinations of prior contexts, embedded words and following contexts. As in the previous chapter, the performance of the ten networks trained for Simulation 4 was investigated by averaging across all the items on which each network was tested. Analyses of activations produced for different test stimuli at different lexical units were carried out as repeated-measures comparisons across these ten computational subjects.

In simulating the results of experiments using fragments of prime sentences it is necessary to decide what position in the sequence of segments presented to the network is equivalent to each probe position tested in the experiments. As in the simulations of priming data reported in Chapter 6, it was assumed that the descriptions of the alignment points given in Chapter 4 provide a reasonable estimate of the phonemic information that is available at each probe position. Thus, AP_1 is placed at the offset of the embedded word with no information about following context available at this point. Although there is evidence from short word responses in gating to suggest that some information from the following word is available at AP_1 , since the network's input is coded as discrete segments, it is not possible to simulate effects of coarticulated information. AP_2 is placed following the onset of the initial segment of the following word – a point at which there will be information that mismatches with the longer competitor for the non garden-path stimuli, but where the garden-path stimuli still match a longer lexical item. AP_3 and AP_4 were placed after the vowel and offset segments of the following syllable.

7.5.1. Recognising embedded words

Figure 7.4a shows the activation of short lexical units in the recurrent network for embedded word stimuli presented in garden-path and non garden-path contexts. These lexical activations are shown alongside equivalent results obtained at each probe position in the priming experiments reported earlier in this chapter (Figure 7.4b). As can be seen by comparing the two graphs, the network overestimates the effect of following context on the activation of short word units. For the non garden-path stimuli, input that mismatches with longer lexical items (at AP_2) leads the network to maximally activate the embedded word. Thus at all probe positions after AP_1 the network produces significantly greater activation for short word units in non garden-path sequences (AP_1 - $t(9)=8.96$, $p<.001$; AP_3 - $t(9)=4.70$, $p<.001$; AP_4 - $t(9)=4.03$, $p<.01$) while the equivalent

comparisons for the priming data are non-significant (and show a numerical trend in the reverse direction).

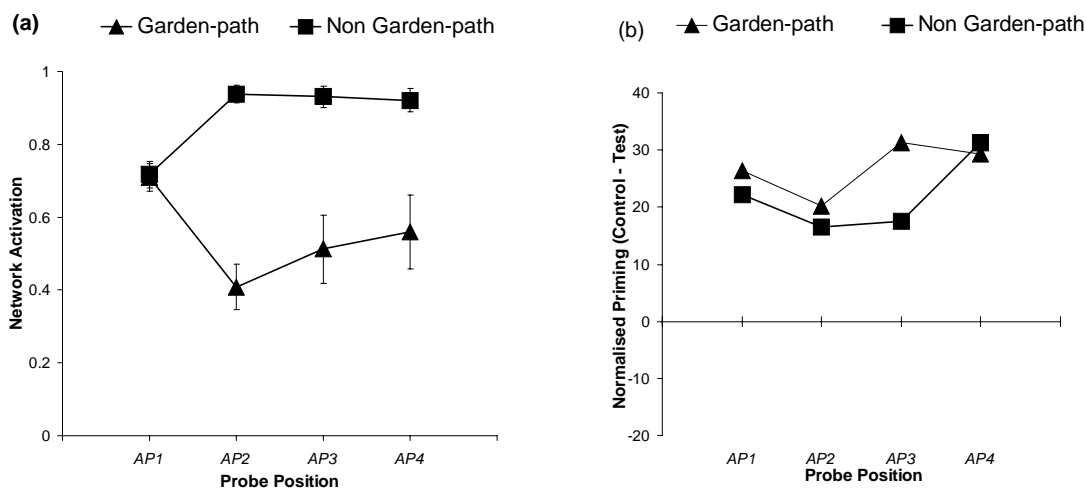


Figure 7.4: Network activation and priming results for short lexical items (CAP) in response to short word stimuli in garden-path (*cap tucked*) and non garden-path contexts (*cap looking*): (a) short word activations over ten networks in Simulation 4. Error bars are one standard error (b) priming results for short targets in Experiments 2 and 4.

This discrepancy between lexical activations observed in the networks and priming data suggests that the model would not predict the null result obtained in comparing the priming of short words in garden-path and non garden-path sequences. Despite the presence of acoustic cues to word length that bias the network towards short word hypotheses at the offset of an embedded word, following context still affects the activation of lexical units for short words. Where following context rules out longer competitors it boosts the activation of short lexical items in the model. No equivalent increase in the priming of short words in non garden-path contexts is observed in the experimental data.

Although the recurrent network simulations reported here lack direct inhibitory connections between lexical units, effects of competition between short and long lexical items are still observed. The comparison between network activations and priming data in Figure 7.4 suggests that even without direct lexical competition, these networks still predict an effect of following context that is not shown in the cross-modal priming data. Explanations of this discrepancy are explored in the concluding section of this chapter.

7.5.2. Ruling out longer competitors

Despite the lack of any significant effect of following context on the activation of short words in the priming data, it is clear that mismatching input is being perceived by listeners. Effects of mismatch were shown by significant differences between the magnitude of priming for long word targets in different following contexts. Greater priming is observed for long word targets from short word stimuli in garden-path contexts, than from stimuli in which there is immediate mismatch between short word stimuli and long lexical items.

The networks from Simulation 4 succeed in simulating this difference between the activation of long words produced by embedded word stimuli with garden-path and non garden-path following contexts. As illustrated in Figure 7.5a, the model predicts significantly greater activation for long words where the following context of an embedded matches a longer competitor compared to sequences where long words are ruled out earlier. This increased activation was significant by pairwise comparisons at all three probe positions where information in the following context was available ($AP_2 - t(9)=15.35, p<.001$; $AP_3 - t(9)=6.09, p<.001$; $AP_4 - t(9)=4.03, p<.01$).

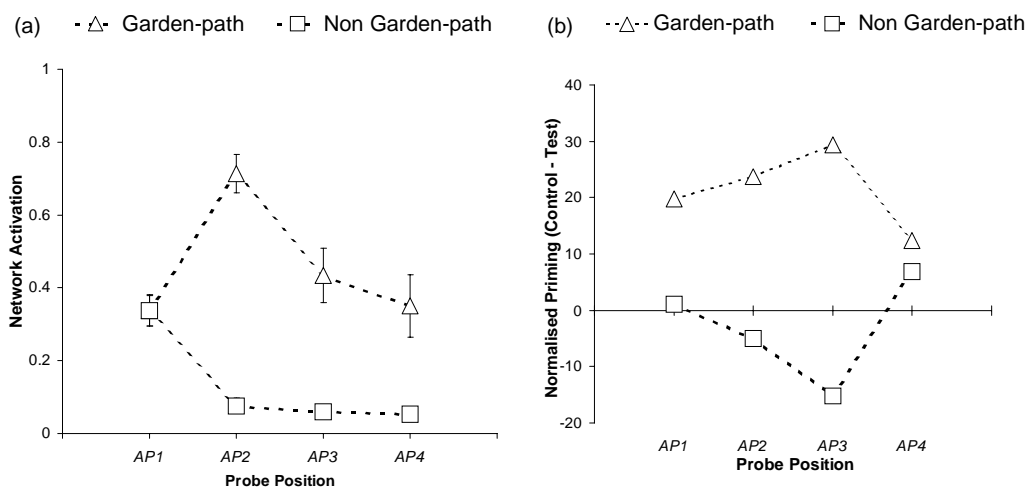


Figure 7.5: Network activation and priming results for long lexical items (CAPTAIN) in response to short word stimuli in garden-path (*cap tucked*) and non garden-path contexts (*cap looking*): (a) long word activations over ten networks in Simulation 4. Error bars are one standard error. (b) priming results for long targets in Experiments 2 and 4.

However, in the priming data shown in Figure 7.5b, differences in priming are only reliable at AP_2 and AP_3 . The numerically greater priming observed for long words from garden-path primes was not significant at AP_1 or at AP_4 . Compared to the priming data, it

seems that the network has greater difficulty in ruling out longer lexical hypotheses for the short word stimuli at the last probe position.

7.6. General discussion

The networks in Simulation 4 are therefore able to account for both the results of Experiment 2 and also the reduced priming of long word targets in Experiment 4. However, some significant discrepancies remain in simulating the priming data for short word targets in Experiment 4. The recurrent networks in Simulation 4 predict that increased priming will be observed for onset-embedded words in contexts where longer words are ruled out immediately after the offset of the embedded word. Since increased priming was not observed in the experiments, it appears that the model has problems accounting for the integration of lexical and acoustic cues to word segmentation. However, before interpreting these results as ruling out the recurrent network account, it is worth considering whether any confounding factors in the experiments or assumptions made in interpreting network activations are responsible for the discrepancies between the networks' predictions and the experimental results.

Alternative explanations of experimental data

Perhaps the simplest argument against interpreting the experimental data as falsifying the model is to suggest that differences between the design of the two sets of experiments are responsible for the inconsistent results. For instance, in Experiment 2 investigating garden-path continuations, prime stimuli came from both short and long words whereas in Experiment 4 prime sentences all contained short words. It is therefore possible that strategic effects produced by different forms of prime-target overlap in the two sets of repetition priming experiments may introduce discrepancies between the priming effects obtained for garden-path and non garden-path stimuli.

In Experiment 4, strategic effects might result from participants noticing that only short word primes are present. These might lead to decreased overall priming of long word targets – explaining the inhibition observed for long word targets at AP_3 . However, this strategic difference would not account for the priming results for short word targets. A strategic explanation would predict increased priming for short words, while this result (predicted by the network through effects of following context) was absent from the

experimental data. Since the lack of significant differences between priming effects for short words in garden-path and non garden-path stimuli is the most difficult result to explain, it seems that strategic differences between the two sets of priming experiments do not help the recurrent network account.

Another concern in interpreting the experimental data is that the comparison of garden-path and non garden-path stimuli does not take place within a single experiment. The conclusion that following context does not affect the activation of short word stimuli is a null result based on a between-experiment comparison. It is therefore possible that the current set of experiments have insufficient statistical power to detect a difference between garden-path and non garden-path stimuli. A within-experiment replication would therefore be valuable in supporting the conclusions drawn from these experiments. In the absence of this data, however, it is necessary to consider whether assumptions made in relating priming data to network activations could account for the failure of the model to simulate the experimental data.

Relating priming data to network activations

In considering how best to interpret this discrepancy between results in the second set of priming experiments and the predictions of models of spoken word recognition, one interesting finding is that in the gating experiment onset-embedded words are more easily identified in non garden-path contexts. Thus, ignoring the overall bias towards short word responses, the pattern of results shown in gating (Figure 7.1) are rather closer to the network's predictions than the priming data. Although not intended to suggest that gating data is to be preferred over results in cross-modal priming, this indicates that behavioural data obtained in psycholinguistic experiments cannot be relied upon to provide a transparent measure of the internal processes of the language processing system.

In comparing network simulations to experimental data, the priming task has not been modelled directly. Instead the simplifying assumption has been made that the activation of lexical units in the network predicts the magnitude of priming observed in more complete simulations that incorporated orthographic inputs and a lexical decision mechanism. However, this assumption may be difficult to justify in the absence of a rather more sophisticated model of language processing. For instance, the recurrent network model investigated by Gaskell & Marslen-Wilson (1997) includes separate output

representations for the phonological form and meaning of the speech input. In simulating the results of semantic and repetition priming experiments, Gaskell and Marslen-Wilson assume that semantic priming is predicted by overlap in the semantic representation, while repetition priming results from similarity in representations of both phonological form and meaning. This account therefore predicts a different relationship between the conditional probability of a word given the current input and the amount of facilitation that is observed of semantically related or repeated targets (Gaskell and Marslen-Wilson, in press). These predictions are confirmed by experiments investigating the relationship between the likelihood of different cohort competitors being present in fragments of speech, and the magnitude of semantic or repetition priming observed.

For semantically related targets, the model predicts that the magnitude of priming is directly proportion to the likelihood of the prime word being present in the input. Where this probability is near zero – for instance, where a much more frequent word also matches the speech input – no reliable priming is predicted. For example, a fragment of speech like /stri:/ would not be predicted to prime a word semantically related to *streak* since this word has a low probability for this fragment (given the presence of a much more frequent cohort competitor *street*). Results of a series of cross-modal priming experiments investigating how the magnitude of semantic priming varies with competitor environment show that the magnitude of semantic priming is directly proportional to the probability of a word given the current input. No priming was observed in these experiments as the conditional probability of the prime word in its cohort environment tends towards zero.

Conversely, in repetition priming, even a word with a low probability of occurrence in its cohort set can be significantly primed through form overlap with its phonological neighbours. So, for the example given above, significant facilitation of responses to the target word *streak* from the spoken fragment /stri:/ would be observed in cross-modal repetition priming. In the model, graphs plotting the magnitude of repetition priming against conditional probability have a significant offset at the origin, as a consequence of the priming effect observed through form-based overlap. These results confirm the predictions of the Gaskell and Marslen-Wilson model since the magnitude of repetition priming, although varying with conditional probability, is not directly related. Form-based priming can arise even where the probability of a word in its cohort set is near zero.

On the basis of these findings, measures of repetition priming may not relate as closely to the activation of lexical/semantic representations as has been assumed in this thesis. Significant repetition priming can be obtained where there is minimal lexical activation of the target word but there is form overlap between the prime and target. Consequently the magnitude of priming effects measured in Experiment 2 and Experiment 4 may not reflect the conditional probability of words in the speech input as directly as was assumed in relating repetition priming data to network simulations. In particular, since form based priming can arise where lexical activations are low, priming effects for short words may be relatively unaffected by the activation of longer competitors.

Since the magnitude of semantic priming has been shown to be more closely correlated with the activations predicted by a probabilistic model (Gaskell & Marslen-Wilson, in press) it may be of interest to use semantic priming rather than repetition priming in follow up experiments. Such data would provide a more stringent test of the recurrent network account developed here in which lexical activations are suggested to be directly proportional to conditional probability.

Lexical activation and probabilistic behaviour

Given these difficulties in relating the predictions of the recurrent network model to behavioural data obtained in priming experiments the status of the probabilistic account of lexical activation presented here remains unclear. The lack of any significant change in priming effects for embedded words depending on whether longer competitors have been ruled out presents a challenge to a probabilistic account. However, it is unclear that these discrepancies would falsify the model rather than simply suggesting that the statistical structure of the network's training environment does not match the properties of natural language.

For instance, the statistical properties of the duration cue used in the network simulations is highly simplified compared to real speech. In order to ensure that the network did not focus on duration at the expense of the segmental input, the duration cue was made deliberately unreliable in the network. Thus, longer words were still robustly activated at the offset of an embedded word, increasing the effect of following context in the network. If the duration cue were made more reliable for the network, then longer competitors

would be more weakly activated initially and following context would play a reduced role in altering the activation of the embedded word.

Secondly, the recurrent network simulations presented in this thesis use a highly impoverished vocabulary by comparison with experimental participants' lexicons. Each embedded word had only one longer competitor in the network, and very few neighbouring lexical items. As a result, only two words were activated during the identification of embedded word stimuli. The network is therefore likely to treat evidence ruling out a longer competitor as ruling in the embedded word. In simulations with multiple long words that all contain the embedded word, this symmetrical competition between short words and a single long competitor would no longer be apparent. Thus networks with a more accurate competitor environment may not predict such large differences between garden-path and non garden-path stimuli.

The failure of the recurrent network account to simulate the experimental data therefore need not indicate that the architecture or computational properties of the system are inappropriate – merely that the training set does not adequately capture the statistical structure of the language. This conclusion is hardly surprising; the training set used for the simulations in this thesis contained only 20 lexical items. However, this finding suggests that caution be exercised in interpreting the failure of the network to exactly simulate the behavioural profile observed in a particular experiment. Small network simulations may be best interpreted as demonstrations of the processing possibilities offered by a particular computational architecture rather than specific predictions about the results of behavioural experiments. In order to simulate behavioural data at an item specific level more realistically sized training sets must be used.