# Processing Semantic Ambiguity: Different Loci for Meanings and Senses

**Jakke Tamminen (jjt2@york.ac.uk)**
**Alexandra A. Cleland (aac5@york.ac.uk)**
**Philip T. Quinlan (ptq1@york.ac.uk)**
**M. Gareth Gaskell (mgg5@york.ac.uk)**
University of York, Department of Psychology
York, YO10 5DD, United Kingdom

**Abstract**

Semantic ambiguity can occur either as a consequence of ambiguity between unrelated meanings (e.g. *bank*) or ambiguity between related senses (e.g. *hook*). Early research did not distinguish between the two, finding that ambiguous words were recognized faster than unambiguous words. More recently it has been shown that words with many meanings suffer from a disadvantage in terms of recognition times, and words with many senses enjoy an advantage over their unambiguous counterparts. We report an auditory lexical decision study in which we apply the Psychological Refractory Period (PRP) logic to investigate the loci of the two types of ambiguity, and argue that they operate at different levels in the word recognition system.

**Keywords:** Psycholinguistics; speech; semantic access

## Introduction

Many words have multiple meanings, that is, they are semantically ambiguous. For example, *bank* can refer equally to a financial institution or to the margin of a river. In such cases the context in which the word occurs can be used to disambiguate the meaning. Evidence has been accumulating in the past decades that, when presented alone, ambiguous words are recognized faster than unambiguous words (see e.g. Rubenstein, Garfield, & Millikan, 1970, for an early demonstration of this with visual lexical decision). Accounts of this effect have relied on the assumption that different meanings of an ambiguous word have separate representations in the lexicon. One such theory was put forward by Jastrzembski (1981), who suggested that, since an ambiguous word has several representations, one of them will be likely to reach a recognition threshold before the single representation of an unambiguous word.

There is, however, some recent evidence suggesting that the issue may be more complicated. While the two meanings of *bank* are clearly semantically unrelated, the majority of ambiguous words have multiple meanings that are closely related. To put it differently, it is common for a word to have many related senses (also referred to as polysemous words, as opposed to homonymous words which have many unrelated meanings). An example of a word with many senses would be *hook*, which can refer to a piece of fishing equipment, as well as to a sharp metal bend, or the act of connecting something.

Most past research into semantic ambiguity has not made a distinction between unrelated meanings and related senses, yet the difference between the two concepts seems important. Thus it is possible that the two types of semantic ambiguity have different consequences for word recognition. Rodd, Gaskell, and Marslen-Wilson (2002) reported visual and auditory lexical decision experiments where they orthogonally manipulated both variables in a 2 x 2 design. Surprisingly, these authors found an ambiguity disadvantage, in that words with many unrelated meanings were recognized more slowly that words with few unrelated meanings. Furthermore, words with many related senses resulted in faster recognition times than words with few senses.

Beretta, Fiorentino, and Poeppel (2005) sought to replicate these findings, and to shed more light on their underlying processes by using magnetoencephalography (MEG). Their behavioral data replicated those of Rodd et al., by showing a senses advantage and a meanings disadvantage. The MEG data showed an interesting effect on the M350 component. This is a component that peaks about 350 ms after stimulus onset, and has been demonstrated to be sensitive to the initial activation of lexical hypotheses. For example, the latency of the M350 has been found to be shorter for high-frequency words than to low-frequency words (Embick, Hackl, Schaeffer, Kelepir, & Marantz, 2001). The M350 was found to be slower to words with many unrelated meanings than few meanings. The opposite pattern was found for senses; words with many senses had shorter M350 latencies than words with few senses.

These results suggest that previous research has confounded the number of unrelated meanings and related senses, and that the ambiguity advantage observed in previous studies may have in fact been caused by items with many senses. Rodd, Gaskell, and Marslen-Wilson (2004) proposed a distributed connectionist model that could explain the senses advantage and the meanings disadvantage. This model explains the meanings disadvantage by postulating separate representations for the different meanings within semantic space. An ambiguous word results in a blend state of activation between the meanings, and the network then needs to move away from the blend state and settle on one meaning. It is this process of competition between the semantic representations that accounts for longer recognition times for ambiguous words.

Related meanings on the other hand have semantic representations that are located close to each other in semantic space. In fact, most of these representations are overlapping, and have developed broad attractor basins. This means that for a word with many senses there is a large area of semantic space that corresponds to that word. Thus, on average, the system will be able to settle on a word with many senses faster than to a word with few senses.

An alternative point of view is to assume that unrelated meanings are represented separately, while words with related senses are represented as a single entry. Beretta et al. (2005) have argued that their findings and those of Rodd et al. (2002) support this view, as the dissociation between effects of meanings and senses seem to indicate that the two types of ambiguity are processed differently. The MEG data strengthen the argument, and show that this difference is reflected even at the earliest stages of word processing.

It is important to highlight the fact that the finding of an ambiguity disadvantage is rare. As discussed earlier, the weight of the evidence so far has favored an ambiguity advantage. Lupker (in press) has suggested that the distinction between unrelated meanings and related senses may be artificial. According to this argument senses which in dictionaries are listed under the same entry, are not always interpreted as being semantically related by participants. The most reliable way of determining the ambiguity status of a word would be to have participants rate words on these dimensions. Taking the controversial nature of these findings into account, it would be valuable to demonstrate the psychological reality of the distinction between senses and meanings in a different paradigm.

The aim of the study reported below was to replicate the reaction time pattern of meanings and senses, and to further investigate the issue of how the two types of ambiguity are represented. For this purpose we employed the Psychological Refractory Period (PRP). In a typical PRP experiment, two tasks (Task 1 and Task 2) are carried out in close succession, and the response times (RTs) to both tasks are monitored as a function of the time interval between the onset of the tasks (stimulus onset asynchrony or SOA) (Pashler, 1994). As the SOA is reduced, RTs to Task 2 become slower, indicating a central "bottleneck" in the simultaneous processing of the two tasks. The slowing down of Task 2 responses is caused by the need for the second task to wait for the bottleneck to finish processing Task 1.

It is generally assumed that the bottleneck corresponds to response-selection processes, while the post-bottleneck phase corresponds to response-execution processes (Pashler, 1994). Hence any effect operating at an early stage of processing should affect the pre-bottleneck stage, and any effect operating at a later, decisional stage should have an influence at the bottleneck. The PRP paradigm can be used to discriminate between the two.

If a variable affecting Task 2 difficulty is manipulated, then the pattern of RTs across SOAs can be used to make inferences about the locus of the variable being

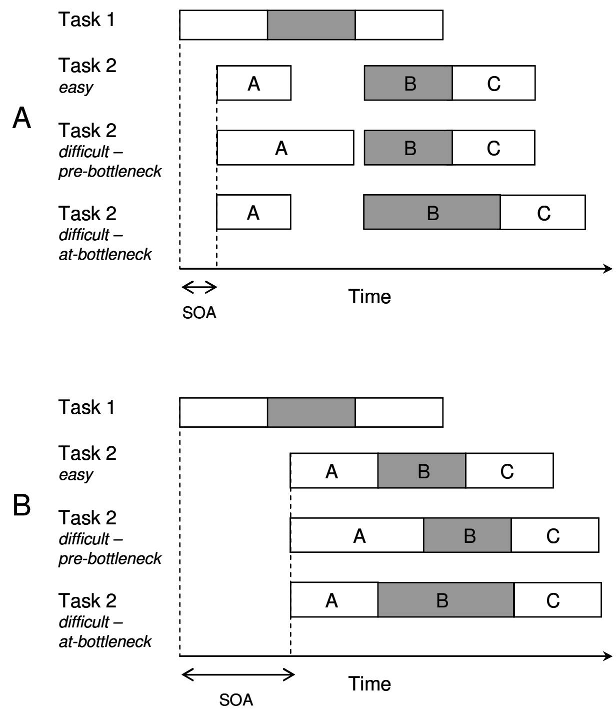


Figure 1: Diagram showing the PRP logic, with a short SOA (A) and a long SOA (B). The grey bars represent the bottleneck. Note that the difficulty manipulation of Task 2 can be located pre- or at-bottleneck, as shown in the figure.

manipulated. At long SOAs the difficulty manipulation should always manifest as faster RTs to the easier condition, as this is equivalent to carrying out two single tasks (see Figure 1, lower panel). Note that in this case the difficulty may affect either pre-, at-, or post-bottleneck processes, the end result is the same in both cases, as shown by the two conditions labeled 'difficult' in the lower panel of the figure.

This RT difference will be apparent at short SOAs only if the process underlying the difficulty manipulation operates at- or post-bottleneck. If it operates pre-bottleneck, it will be able to take advantage of the slack time created while waiting for the bottleneck to clear, and no RT difference between the conditions will result, as shown in the upper panel of Figure 1. The former pattern, where the difficulty manipulation is apparent across all SOAs, is termed *additive*, and the latter pattern, where the difficulty effect disappears at the short SOAs, is termed *underadditive*.

There is some recent work applying the PRP logic to spoken word recognition. Variables examined in these studies have shown underadditive patterns. Cleland, Gaskell, Quinlan, and Tamminen (2006) found a word frequency effect that was underadditive with SOA, in both auditory and visual lexical decision (although cf. McCann, Remington, & Van Selst, 2000). Tamminen, Cleland, Quinlan, and Gaskell (submitted) found an underadditive pattern when manipulating subphonemic mismatches in a phonemic decision task. These studies suggest that variables affecting spoken word recognition typically occur pre-bottleneck.

If the two types of semantic ambiguity discussed above are represented or processed at different loci with respect to the bottleneck, it should be possible to observe different RT patterns for the two classes of stimuli across SOAs, provided that one requires processing pre-bottleneck and the other at-bottleneck. It is worth noting that this paradigm cannot distinguish between at- and post-bottleneck effects. However, this is not a crucial distinction here, as there is little reason to believe that a semantic manipulation would affect response-execution. In the study reported here, we apply the experiment of Rodd et al. (2002) to the PRP paradigm, with a binary color judgment task used as Task 1, and an auditory lexical decision as Task 2.

## Method

**Participants** Thirty-two participants from the University of York were recruited. Mean age was 20 (range 18-22). Twenty-four were female, 8 male. Thirty-one were right-handed, one was left-handed. All participants were native English speakers, and none reported any visual, hearing, or speech disorders. Participants received either cash payment or course credit.

**Stimuli and materials** The visual stimuli included two shapes, a circle and a square. Both shapes had a blue, green, and an unfilled version. The shapes were bitmap images, measuring 8 cm in width and in height on the screen.

The words used were taken from Experiment 3 of Rodd et al. (2002). In order to divide the items equally across two stimulus alignment conditions, one item from each of Rodd et al.'s condition list was removed, resulting in 22 words in each of four semantic conditions (many meanings with many senses, many meanings with few senses, few meanings with many senses, and few meanings with few senses). The words were matched in CELEX (Baayen, Piepenbrock, & Van Rijn, 1993) frequency (log-transformed), number of phonemes, uniqueness point, concreteness, and familiarity.

These items were recorded by a male native English speaker in a sound-attenuated booth, using a Sennheiser ME40 microphone and Pioneer PDR 509 recording system. The sound files were normalized in wav format (mono, 44 kHz sample rate, with 16 bit resolution). The mean duration of the recordings was 466 ms for many meanings with many senses, 498 ms for many meanings with few senses, 443 ms for few meanings with many senses, and 477 ms for few meanings with few senses.

In addition to the experimental items, 88 filler words and 176 nonwords were used. The filler words were taken from a pool of items that contained high-frequency words (less than ten occurrences per million) and low-frequency words (more than 100 occurrences per million). Half of the fillers were taken from the high and the rest from the low frequency groups. The nonwords were based on real words with one phoneme changed (initial, middle, or last phoneme of the word) to make a pronounceable nonword. The filler words and nonwords were recorded using the same speaker and equipment as above, but were recorded during a different session.

**Design** The experiment had eight conditions, defined by two levels of stimulus alignment (delay between the onset of the colored shape and the onset of the final phoneme of the spoken word could be 100 or 1000 ms, henceforth referred to as stimulus asynchrony) and four levels of the semantic variable (many meanings with many senses, many meanings with few senses, few meanings with many senses, and few meanings with few senses).

Figure 2 is a graphical representation of an experimental trial. The stimulus alignment used in this experiment differs somewhat from typical PRP experiments where the stimuli are usually aligned from the onset of Task 1 stimulus to the onset of Task 2 stimulus. However, unlike written words, the informational content of spoken words unfolds over time. Thus it would be inappropriate to align the stimuli relative to the onset of the spoken word, as the word cannot be identified until its uniqueness point, which tends to be towards the end of the word. We chose the onset of the final phoneme burst as the stimulus alignment point here, as that more accurately represents the point in time where the spoken word can be identified (cf. Cleland et al., 2006).

The aligning of the auditory stimulus with respect to its final phoneme burst creates an additional complexity. In PRP experiments it is important that the Task 1 stimulus begins before the Task 2 stimulus. This helps to ensure that Task 1 is processed first. With auditory words, when the stimulus asynchrony is short, this is not the case. As shown in Figure 2, the onset of the word extends beyond the onset of the colored shape. To deal with this problem it was necessary to extend the Task 1 stimulus. We did this by adding another shape to precede the colored shape.

All items were rotated through the stimulus asynchrony conditions so that, across participants, each item occurred with both the long and the short asynchronies, and also so that each item occurred with both a blue and a green colored shape.

**Procedure** Participants were informed that they had two tasks to carry out, a color discrimination task (Task 1), and a lexical decision task (Task 2). They were asked to respond as quickly as possible to both tasks, but to give emphasis to the color task. Each trial started with a fixation cross (+) presented on the screen for 1000 ms. This was replaced by the sequence of two shapes, one of which was unfilled, and the other one was colored. In half of the trials the first shape was colored, in the other half the second shape was colored. Only filler items were used in the trials where the first shape was colored. All experimental items occurred in the condition where the colored shape was the second one, thus ensuring that they were all correctly aligned to the shapes. Each shape stayed on screen for 375 ms. The word or nonword was played through headphones at a time

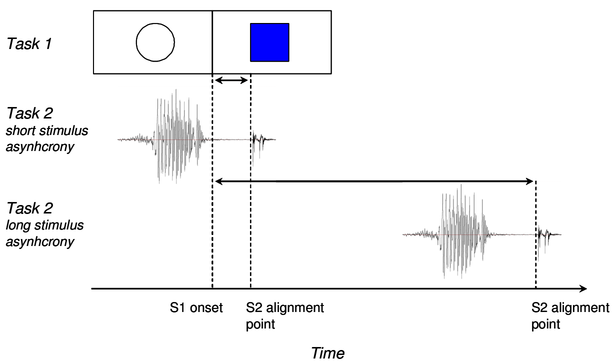



Figure 2: An example of a trial illustrating the two stimulus asynchrony conditions.

determined by the stimulus asynchrony condition. The trial ended after two responses were made or after 3750 ms had elapsed from the onset of the first shape. Response times to Task 1 were measured from the onset of the colored shape, and in Task 2 from the onset of the word/nonword.

A practice block of 32 (16 words, 16 nonwords) items preceded the experimental trials. E-prime was used for stimulus presentation and timing on a PC running on a 1.84 GHz processor. The order of items was randomized by the software for each participant, and a rest break was given half way through the experimental trials. Visual stimuli were presented on a Sharp 17" flat panel TFT monitor, and responses were recorded from a Cedrus response box. The response box was placed so that the participants always made Task 1 responses with their left hand and Task 2 responses with their right hand.

## Results and Discussion

Trials where an error was made either in Task 1, Task 2, or both tasks were excluded from the RT analysis. This resulted in the exclusion of 11% of trials. In addition to this, all responses where the reaction time was above 3000 ms or below 100 ms were excluded. To reduce the effects of remaining outliers, the RT data were subjected to an inverse transformation (Ulrich & Miller, 1994).

**Errors** No effects of semantic variables, stimulus asynchrony, or interactions were found in the percentages of errors made to either Task 1 or Task 2, (all *p*s > .05).

**Task 1** Mean RTs to Task 1 are presented in Table 1. A repeated measures ANOVA with stimulus asynchrony, number of senses, and number of meanings as factors was carried out on the data. No main effects or interactions showed significant results (all *p*s > .05).

**Task 2** An analysis of the recordings of the experimental items revealed that recordings of words with few senses tended to be of longer duration than words with many senses. The duration from the onset of the word to the onset of the final phoneme burst was on average 35 ms longer in words with few senses. To ensure the reliability of the data reported below, RTs were measured from the onset of the words instead of the alignment point. This is a more standard way of measuring word recognition times, and, as one cannot know for sure when the critical information identifying a words comes in, is more robust against a potential confound.

Table 1: Mean RTs (ms) in each semantic and stimulus asynchrony condition. T1 = Task 1, T2 = Task 2.

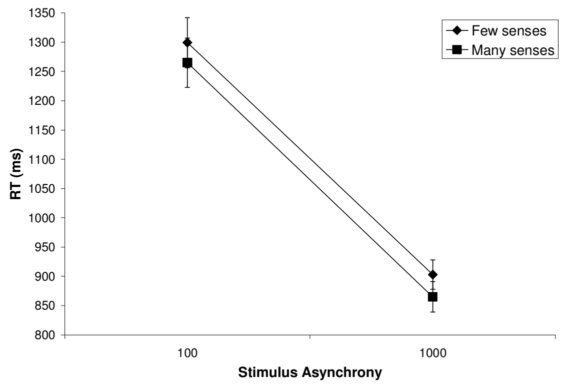
| | | Stimulus asynchrony | | | |
|---|---|---|---|---|---|
| | | 100 | | 1000 | |
| Meanings | Senses | T1 | T2 | T1 | T2 |
| Many | Few | 575 | 1282 | 558 | 911 |
| Many | Many | 568 | 1252 | 559 | 895 |
| Few | Few | 586 | 1315 | 531 | 894 |
| Few | Many | 571 | 1278 | 536 | 835 |

The data for the main effects are presented in Figure 3. A main effect of stimulus asynchrony was found, $F_1(1, 31) = 231.44$, $p < .001$, $F_2(1, 84) = 697.81$, $p < .001$, showing that participants were on average 398 ms faster to make a response in the long stimulus asynchrony condition. A reliable main effect of number of senses was also apparent (by-items analysis is marginally significant), $F_1(1, 31) = 10.10$, $p < .01$, $F_2(1, 84) = 3.75$, $p = .056$, reflecting longer RTs to words with few senses, compared with words with many senses (36 ms). The main effect of number of meanings did not reach significance, $F_1(1, 31) = 0.33$, $p > .05$, $F_2(1, 84) = 0.02$, $p > .05$. Stimulus asynchrony interacted reliably with number of meanings, $F_1(1, 31) = 6.16$, $p < .05$, $F_2(1, 84) = 5.06$, $p < .05$. No other interactions, including the interaction between senses and stimulus asynchrony, were significant (*p*s > .05).

Simple planned comparisons were carried out to confirm the nature of the RT patterns. The effect of meanings was statistically significant at the long asynchrony, $F_1(1, 31) = 5.74$, $p < .05$, $F_2(1, 84) = 4.19$, $p < .05$, where there was a 38 ms advantage for words with few unrelated meanings. The effect was not significant at the short asynchrony, $F_1(1, 31) = 3.61$, $p > .05$, $F_2(1, 84) = 1.22$, $p > .05$, where there was a numerical advantage of 29 ms for words with many unrelated meanings.

The effect of senses was significant at the long asynchrony, $F_1(1, 31) = 11.02$, $p < .01$, $F_2(1, 84) = 4.34$, $p < .05$. Participants were 38 ms faster to respond to words with many senses than to words with few senses. The effect approached significance by-participants at the short asynchrony, $F_1(1, 31) = 3.51$, $p = .07$, $F_2(1, 84) = 1.43$, $p > .05$, where responses to words with many senses were 34 ms faster than to words with few senses.

There is a clear dissociation between the RT patterns to the two types of ambiguity. Words with many senses had an advantage over words with few senses, and this effect was additive with stimulus asynchrony, which suggests that the effect is taking place at- or post-bottleneck. The pattern is a)




b)

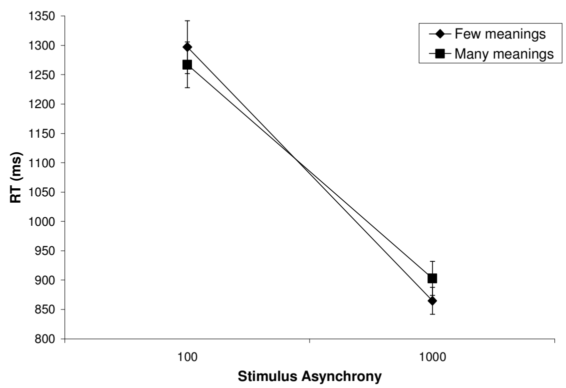



Figure 3: Mean RTs, measured from word onset, to words with few and many senses (a), and to words with few and many unrelated meanings (b). Error bars represent standard error.

different for the unrelated meanings manipulation. Here the effect is underadditive with stimulus asynchrony, a pattern that can be taken to mean that the effect is taking place pre-bottleneck.

The pattern of RTs for the number of senses manipulation also indicates that the effects observed could not be attributed to the fact that the few senses words were of a slightly longer duration than the many senses words. At the short stimulus asynchrony, and for both classes of words, the majority of the spoken item will have unfolded prior to the presentation of the colored shape. At some point after this Task 1 will demand bottleneck processes and this will eventuate in some amount of slack time before the word can engage central processes. On the understanding that stimulus encoding processes associated with the word occur at a pre-bottleneck stage, then any difference in the duration of these that are due to the differences in utterance length will be soaked up in the slack time produced by Task 1. However, at longer asynchronies in which Task 1 stimulus precedes the presentation of Task 2 stimulus, any effects due to difference in utterance length become more apparent. In other words a classic underadditive pattern ought to obtain if the effects are indicative of differences in utterance duration. As can be seen from Figure 3, quite a contrary pattern is shown: the senses effect is additive with stimulus asynchrony.

## General Discussion

Some authors (e.g. Lupker, in press) have recently questioned the psychological validity of the distinction between the two types of semantic ambiguity discussed here; ambiguity between unrelated meanings and ambiguity between related senses. Many studies have relied on dictionary definitions to assign words into the semantic ambiguity conditions. However, the way lexicographers see the distinction may not correspond with judgments made by people, or with the cognitive architecture of language representation. Two aspects of our data should alleviate these concerns. Firstly, we provide a replication of the dissociation of the two types of ambiguity in terms of RTs. Secondly, the finding that senses and meanings affect processing at different loci with respect to a central attentional bottleneck clearly points to two separate processes.

As stated, our data successfully replicate the RT pattern of previous studies examining number of senses and meanings. Recall that the long stimulus asynchrony condition is equivalent to carrying out two single tasks. Thus this condition is the closest equivalent to the Rodd et al. (2002) and Beretta et al. (2005) studies, which found faster RTs to words with many senses than to words with few senses, and slower RTs to words with many meanings than to words with few meanings. The same pattern was found in our study.

Another purpose of the study was to distinguish between the two types of semantic ambiguity. The PRP logic allows us to make inferences about the nature of processing of the two (Pashler, 1994). Different RT patterns were found for senses and unrelated meanings. Participants were slower to respond to words with few senses than to words with many senses at long and short asynchronies, suggesting that the process underlying this effect was not able to take advantage of the cognitive slack time created while waiting for Task 1 to clear the bottleneck. This is taken as evidence that the process is taking place at- or post-bottleneck. This additive pattern stands in contrast to the underadditive pattern discovered in the case of unrelated meanings. Participants were slower to respond to words with many unrelated meanings than to words with few meanings at the long asynchrony. This ambiguity disadvantage disappeared at the short asynchrony, indicating that the processing of this type of ambiguity is able to take advantage of the slack time, meaning that it must be processed pre-bottleneck. While the ambiguity between unrelated meanings is being resolved at an early stage, the ambiguity between related senses is resolved at a later stage.

Due to the distinction between early and late stages of processing afforded by the PRP logic, we are also able to address the question of how semantic ambiguity is processed in the two cases, and to evaluate existing theories. Rodd et al. (2004) proposed a connectionist model with distributed semantic representations to explain the meanings disadvantage and the senses advantage, as described in the Introduction. Our current data provide a significant challenge to this model. On the face of it, it is difficult to see how such a model could accommodate both pre- and at-bottleneck effects within the same representational level.

One explanation would be to propose that unrelated meanings are processed pre-semantically, and related senses semantically. As seen in the Introduction, the original ambiguity advantage was interpreted in the framework of separate representations for each meaning of an ambiguous word (Jastrzembski, 1981). If each unrelated meaning has its own pre-semantic lexical representation, then it becomes possible to explain the meanings disadvantage through the operation of lexical competition. Orthographic or phonological input would activate the representations of all the meanings, which would then engage in competition, thus slowing recognition. Furthermore, the frequency of each individual meaning is likely to be lower than that of a word with one unambiguous meaning, providing another mechanism through which the effect may operate.

Our data suggest that the senses effect on the other hand is a late occurring effect. This would fit in well with a semantic process. Rodd et al. (2002) discuss the possibility that words with many and few senses differ in the amount of semantic information they contain. Words with many senses would be rich in semantic features, an advantage which can lead to more stable representations, and faster settling times in distributed networks. The mechanisms proposed above seem to accommodate our findings; competition between pre-semantic word representations would lead to an ambiguity disadvantage and would likely to be a pre-bottleneck effect. A senses advantage operating at the level of semantic features would result in an ambiguity advantage, and take place at the bottleneck.

## Acknowledgments

The research was supported by BBSRC research grant S19172 awarded to G.G. and P.Q.

## References

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM).* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Beretta, A., Fiorentina, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access. *Cognitive Brain Research, 24,* 57-65.

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 104-119.

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M, & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research, 10,* 345-348.

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13,* 278-305.

Lupker, S. (in press). Representation and processing of lexically ambiguous words. To appear in M. G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics.* Oxford: Oxford University Press.

McCann, R. S., Remington, R. W., & Van Selst, M. (2000). A dual-task investigation of automaticity in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1352-1370.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220-244.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46,* 245-266.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28,* 89-104.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9,* 487-494.

Tamminen, J., Cleland, A. A., Quinlan, P. T., & Gaskell, M. G. (submitted). The nature of phoneme resolution in spoken word recognition.