This is a repository copy of *Disambiguating the ambiguity disadvantage effect: Behavioral and electrophysiological evidence for semantic competition*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/158320/

Version: Accepted Version

Article:

Maciejewski, G and Klepousniotou, E orcid.org/0000-0002-2318-0951 (2020) Disambiguating the ambiguity disadvantage effect: Behavioral and electrophysiological evidence for semantic competition. Journal of Experimental Psychology: Learning, Memory, and Cognition. ISSN 0278-7393

https://doi.org/10.1037/xlm0000842

© 2020 American Psychological Association. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: https://psycnet.apa.org/doi/10.1037/xlm0000842

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Running head: AMBIGUITY DISADVANTAGE

Disambiguating the ambiguity disadvantage effect: Behavioral and electrophysiological evidence for semantic competition

Greg Maciejewski[1,2] & Ekaterini Klepousniotou[1]

[1] School of Psychology, University of Leeds, UK

[2] School of Education and Social Sciences, University of the West of Scotland, UK

Corresponding author:

Dr. Ekaterini Klepousniotou

School of Psychology

University of Leeds

Leeds

LS2 9JT

Phone: +44 (0)113 3435716

Email: e.klepousniotou@leeds.ac.uk

Abstract

Semantic ambiguity has been shown to slow comprehension, though it is unclear whether this “ambiguity disadvantage” is due to competition in semantic activation or difficulties in response selection. We tested the two accounts by examining semantic relatedness decisions to homonyms, or words with multiple unrelated meanings (e.g., “football/electric fan”). Our behavioral results showed that the ambiguity disadvantage arises only when the different meanings of words are of comparable frequency, and are thus activated in parallel. Critically, this effect was observed regardless of response-selection difficulties, both when the different meanings triggered inconsistent responses on related trials (e.g., “fan-breeze”) and consistent responses on unrelated trials (e.g., “fan-snake”). Our electrophysiological results confirmed that this effect arises during semantic activation of the ambiguous word, indexed by the N400, not during response selection. Overall, the findings show that ambiguity resolution involves semantic competition and delineate why and when this competition arises.

Keywords: lexical/semantic ambiguity; homonymy; meaning frequency; semantic processing; N400

1 Introduction

The vast majority of the words we use are in some way ambiguous, hence the ability to select a single, contextually appropriate meaning without being overtly distracted by the myriad of other possible meanings is a crucial component of any theory of language comprehension. Indeed, the importance of understanding how multiple meanings are represented and accessed is highlighted by the extensive literature dedicated to this issue over the past few decades (for a recent review, see Rodd, 2018).

One unclear finding in this literature is that of slower response/reading times for ambiguous versus unambiguous words in tasks that require meaning selection in neutral context or isolation. This so-called "ambiguity disadvantage" effect has been typically observed for homonyms, words with multiple unrelated meanings, either in late-disambiguation sentence reading (e.g., "He found the coach was too hot to sleep in"; Duffy, Morris, & Rayner, 1988; Frazier & Rayner, 1990; Rayner & Duffy, 1986) or semantic relatedness decisions (e.g., "hide-conceal/skin"; Gottlob, Goldinger, Stone, & Van Orden, 1999; Hoffman & Woollams, 2015; Pexman, Hino, & Lupker, 2004; Piercey & Joordens, 2000). Although the effect appears to be robust, little is still known as to why it arises, and what it reveals about the representations and processes involved in ambiguity resolution. Here, we focus on homonyms and examine two prominent accounts of the effect – semantic competition and decision making.

The ambiguity disadvantage is an inherent prediction of the "distributed" view of lexical-semantic representation (for an overview, see Seidenberg, 2007). In short, connectionist models postulate that words are represented by units corresponding to their orthographic and semantic features. These units are distributed, in the sense that a single unit contributes to the representation of multiple words that share the same feature. There are connections among the orthographic and semantic units which, as a result of learning, acquire different weights reflecting the appropriate form-to-meaning mapping. Thus, within this framework, it is the weights on the connections that determine the ease of semantic activation.

For unambiguous words, the orthographic pattern of activation is always associated with the same semantic pattern, which strengthens the connections and facilitates future form-to-meaning mapping. Ambiguity, on the contrary, precludes such a benefit. The orthographic pattern for words such as “bank” is ambiguous and gives rise to a “blend state”, or partial activation of the different semantic representations (“money/river bank”). As semantic activation increases, the representations begin to compete for full activation, as only one of them can be activated to complete the disambiguation process. According to connectionist models of ambiguity (e.g., Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd, Gaskell, & Marslen-Wilson, 2004), it is this semantic competition, due to multiple form-to-meaning mappings, that may account for the ambiguity disadvantage in word comprehension.

The semantic competition account proposed in the connectionist models has been challenged by Pexman et al. (2004), who argued that the ambiguity disadvantage is due to decision-making difficulties in response selection. Their semantic relatedness decision tasks revealed a substantial processing cost for ambiguous words on related trials, regardless of which meaning the targets instantiated (e.g., “hide-conceal/skin”). Interestingly, there was no such cost on unrelated trials, where the same words were paired with unrelated targets (e.g., “hide-glass”). Pexman et al. (2004) reasoned that if the ambiguity disadvantage were due to semantic activation processes, its effects would be observed on both related and unrelated trials. On related trials, participants need to resolve ambiguity because a blend state is not sufficient to support a correct response (e.g., only one of the meanings of “hide” is related to “conceal”). On unrelated trials, participants may not need to resolve ambiguity (e.g., both meanings of “hide” are unrelated to “glass”), but their response is still entirely based on semantic activation. To accommodate their findings, Pexman et al. (2004) posited that the processing cost specific to related trials may be a task artefact caused by response conflict. Since the different meanings of homonyms are inconsistent with the same response to a related target, the cost may reflect the need to decide on which meaning should serve as response input. Critically, no such response conflict arises on unrelated trials, hence the null ambiguity effect when making relatedness decisions to unrelated word pairs.

Further support for the idea that the ambiguity disadvantage lies in decision-making difficulties comes from Hino, Pexman, and Lupker’s (2006) semantic categorisation studies. Since the different meanings of ambiguous words often fall into different categories (e.g., “crane” in reference to the living/non-living category), and may therefore create response conflict similar to that in relatedness decision tasks, the researchers focused on “no” trials where neither meaning fell into a category in question (e.g., “bear” in reference to the vegetable category). Their results showed a processing cost for homonyms when the task involved broad living-object or human-related categories, but not when it involved narrow animal or vegetable categories. Hino et al. (2006) attributed this pattern of responses to the nature of the decision category (see also Hargreaves, Pexman, Pittman, & Goodyear, 2011). When the category is broad, participants must retrieve a large number of semantic features of the target word’s referent and decide whether any of them is true of the category. For ambiguous words, this may take considerably longer because participants need to retrieve and analyze features of multiple word referents, in case one of them falls into the category in question. In contrast, when the category is well-defined and narrow, participants may be able to respond based on a small number of features that are likely true of all the word referents, whilst ignoring irrelevant features that would otherwise slow processing. The overall argument, then, is that the ambiguity disadvantage arises only when task-relevant decisions are somewhat more difficult to make.

In summary, the challenge of explaining the ambiguity disadvantage has provided a strong impetus to the development of different accounts of ambiguity representation. Under the semantic competition account (Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004), the delay in comprehension arises because ambiguous words, in particular homonyms, have separate semantic representations that compete for activation. Under the decision-making account (Hino et al., 2006; Pexman et al., 2004), the delay arises due to task-specific response-selection demands. The semantic representations are assumed to be activated independently, without giving rise to any interference. Pexman et al. (2004) argued that such an explanation would hold true if we assumed that the different meanings of ambiguous words are represented in separate subsets of semantic memory, such that, for example, the institution-related meaning of the word “bank” is represented within one semantic space, whereas the river-related meaning is represented within another.

Overall, then, there is a clear need to establish the locus of the ambiguity disadvantage before we make any further inferences about the structure of the mental lexicon.

The present study directly tested the semantic competition and decision-making accounts by investigating the ambiguity disadvantage in semantic relatedness decision tasks, in which ambiguous/unambiguous words were followed by related/unrelated targets. For ambiguous words, we focused on homonyms, expecting that if any form of ambiguity produced competition at the semantic level, it would be, foremost, observed for homonyms whose unrelated meanings are unanimously assumed to have separate semantic representations (for a review, see Eddington & Tokowicz, 2015). In Experiment 1, we contrasted homonymous and non-homonymous words on related and unrelated trials to replicate the ambiguity disadvantage in the first place. In Experiment 2, we used EEG measurements to determine when this effect is in play – in other words, whether it arises during the processing of the ambiguous word itself, as predicted by the semantic competition account, or during response selection upon the presentation of the target, as suggested by the decision-making account.

2 Experiment 1

Unlike previous relatedness decision studies (Gottlob et al., 1990; Pexman et al., 2004; Piercey & Joordens, 2000), Experiment 1 made a clear distinction between homonyms with balanced (e.g., “football/electric fan”) and unbalanced meaning frequencies (e.g., “blue/spacious pen”). The rationale was that although all meanings seem to be activated upon reading an ambiguous word, the broader literature on ambiguity resolution indicates that the level and time-course of this activation are influenced by meaning frequency, or dominance (for a review, see Twilley & Dixon, 2000). In particular, when ambiguous words are encountered in isolation or neutral context (as in the present study), readers are biased towards the high-frequency (HF) meaning, such that activation of the low-frequency (LF) counterpart is noticeably weaker, delayed, and transient (Frost & Bentin, 1992; Klepousniotou, Pike, Steinhauer, & Gracco, 2012; Simpson & Burgess, 1985). Drawing on this line of research, we argue that any adequate account of how activation of multiple meanings affects word comprehension must take into account the role of meaning frequency in the activation process, or the distinction between balanced and unbalanced homonyms. For example, if semantic competition does arise, one would expect it to be maximal for balanced homonyms whose different meanings are initially activated to the same extent and in parallel (in neutral or out of context). For unbalanced homonyms, on the other hand, the impact of meaning frequency should eliminate, or at least reduce, the competition. Readers may fully retrieve and select the HF meaning very fast, such that the LF counterpart does not reach a sufficient level of activation to engage in the competition.

The idea that meaning frequency modulates ambiguity effects in word processing is not entirely new, as there have been a few studies that either controlled for (Armstrong & Plaut, 2016; Mirman, Strauss, Dixon, & Magnuson, 2010) or manipulated this property (Brocher, Koenig, Mauner, & Foraker, 2018; Grindrod, Garnett, Malyutina, & den Ouden, 2014; Klepousniotou & Baum, 2007; Klepousniotou et al., 2012; MacGregor, Bouwsema, & Klepousniotou, 2015). In particular, Armstrong, Tokowicz, and Plaut (2012) suggested that the impact of homonymy in word recognition depends on the relative frequencies of the multiple meanings, such that there is a slight slowing in lexical decisions to balanced but not unbalanced homonyms (but cf. Grindrod et al., 2014; Klepousniotou & Baum, 2007). As for the impact in word comprehension, late-disambiguation sentence-reading studies (Duffy et al., 1988; Rayner & Duffy, 1986) reported a similar pattern of results - a processing cost for balanced but not unbalanced homonyms. This is in line with Kawamoto's (1993) model simulations predicting the ambiguity disadvantage to be more pronounced when the different meanings are of comparable frequency, and thus equal competitors in the race for activation.

Taken together, we sought to replicate and further examine the ambiguity disadvantage for balanced but not unbalanced homonyms in semantic relatedness decisions. This was necessary for Experiment 2 where we used EEG measurements to establish when and why the disadvantage arises, separating early semantic activation processes during prime presentation from late response-selection processes during target presentation. Note that Experiment 1 involved two versions that differed in the duration of the ambiguous/unambiguous prime (200 ms in Experiment 1a, 700 ms in Experiment 1b). The rationale was that although studies indicated a delay in LF-meaning activation for homonyms on the whole (Frost & Bentin, 1992; Simpson & Burgess, 1985; Simpson & Krueger, 1891), it is unclear how substantial this delay might be for highly unbalanced homonyms, such as those used in the present study. Extending the prime duration was therefore essential to confirm that unbalanced homonymy does not produce the ambiguity disadvantage, either due to semantic competition or decision making, even when there is enough time to retrieve the LF meaning.

2.1 Method

2.1.1 Participants

Participants were University of Leeds students and staff who participated for course credit or £3. There were 35 participants [30 females, aged 19-35 (M = 25.8, SD = 4.9)] in Experiment 1a and a different group of 30 [21 females, aged 18-42 (M = 21.3, SD = 5.5)] in Experiment 1b. All participants were monolingual native speakers of British English with no known history of language-/vision-related difficulties or disorders. They were right-handed, as confirmed with the Briggs-Nebes (1975) modified version of Annett's (1967) handedness inventory. The experiment received ethical approval from the School of Psychology, University of Leeds Ethics Committee.

2.1.2 Stimuli

Prime words were 28 balanced homonyms (e.g., "fan), 28 unbalanced homonyms (e.g., "pen"), and 56 non-homonyms (e.g., "crew") that were split into two sets (1 & 2). All homonyms were selected from the British norms of meaning frequency (Maciejewski & Klepousniotou, 2016). The four sets of primes were statistically comparable (all Fs < 1) with respect to 14 lexical and semantic variables, such as form frequency and semantic diversity. For more information on prime-word selection and matching, see Section 1.1 in the Supplementary Material.

We paired each homonymous prime with four target words: two semantically related and two semantically unrelated. For unbalanced homonyms, one target related to the HF meaning of the prime (e.g., “pen-ink”), while the other related to the LF meaning (e.g., “pen-farmer”). The same manipulation was used for balanced homonyms, although the difference in meaning frequencies for these items was much smaller, as evident in the norms (Maciejewski & Klepousniotou, 2016). For non-homonyms, the two related targets (A & B) referred to the same interpretation of the prime (e.g., “fake-truth/fraud”). We also paired all primes with two targets (A & B) that were unrelated to either of their meanings (e.g., “fan-snake/cancel”). This aimed to equalize the number of related and unrelated word pairs in the experiment (all listed in the Appendix). The 16 sets of targets were statistically comparable (all Fs < 1) with respect to 14 lexical and semantic variables, such as form frequency and semantic diversity. For more information on target-word selection, matching, and prime-target relatedness pre-test, see Sections 1.2 and 1.3 in the Supplementary Material. For examples of different prime-target word pairs, see Table 1 below.

>> Insert Table 1 here <<

2.1.3 Procedure

The relatedness decision task was programmed in EPrime 2.0 (Schneider, Eschman, & Zuccolotto, 2010). The task was to decide whether the prime and the target were related in meaning by pressing a keyboard button. Participants pressed the L button for “related” with their dominant (right) hand or the A button with their left hand. Both response speed and accuracy were equally emphasized in the instructions, and participants were instructed and given examples as to what constitutes semantic relatedness.

The stimuli were pseudo-randomly divided into four blocks, such that each block contained the same number of trials in the different conditions. Participants responded to each prime four times, but none of the primes appeared more than once within the same block. The order of blocks was counter-balanced across participants. The order of trials within each block was pseudo-randomized, such that no more than three related/unrelated word pairs appeared consecutively. The task began with a practice block comprising two examples of each condition (N = 32) and feedback on each response. There were two one-minute breaks - one after the practice block and the other after the first two experimental blocks. Following each break, participants first responded to eight filler trials (not included in the analysis) that aimed to help them get back to the habit of quick responding.

In Experiment 1a, trials began with a 500 ms fixation cross. After 100 ms, the prime and the target were presented for 200 ms and 500 ms, respectively, with a 50 ms interval in between. Once the target disappeared, there was 1500 ms for response execution followed by a 100 ms inter-trial interval. Participants could make relatedness decisions as soon as the target appeared, but they had to respond within the first 1500 ms (i.e., responses of 1500-2000 ms were deemed too slow and would be excluded from analyses). In Experiment 1b, the only difference was the longer prime-word duration (700 ms instead of 200 ms).

2.2 Results

In Experiment 1a, two of the 35 participants were removed from analyses – one due to a large number of errors on related trials (63.8%) and the other due to slow and variable responses (M = 899.9, SD = 182.9). In Experiment 1b, one of the 30 participants was removed due to a large number of errors on related trials (54.5%). In Experiment 1a, we also removed the four targets of one of the non-homonyms as these were inadvertently paired with a different prime. For RTs, we excluded errors (19% and 17% of trials in Experiments 1a and 1b, respectively) and any responses that were two SDs above/below a participant's mean in a given condition (4% and 3% of trials in Experiments 1a and 1b, respectively). The remaining RTs were log-transformed to further minimize the impact of potential outliers and to normalize the residual distribution[1].

Accuracy and latency data were analyzed using logit/linear mixed-effects models with the factors of Prime (balanced homonym, unbalanced homonym, non-homonym$_1$, non-homonym$_2$) and Target (HF-meaning/A, LF-meaning/B). RT models also included Block (1, 2, 3, 4), though effects involving this factor are not reported because its sole purpose was to account for potential practice or prime-repetition effects, and no such effects were detected[2] (Pollatsek & Well, 1995). Terms involving Block were removed from accuracy models due to non-convergence. Related and unrelated trials were analyzed separately, as our preliminary analyses showed a significant effect of Trial (i.e., slower but more accurate responses on unrelated trials) that always interacted with the effects of Prime and Target. These preliminary analyses were otherwise the same as those reported below. Furthermore, we were concerned that contrasts between ambiguity effects on related and unrelated trials would not be readily interpretable, as it is not entirely clear what processes underlie performance on unrelated trials, and whether/how they differ from those on related trials.

Each model included significant random intercepts for subjects and items. Following Barr, Levy, Scheepers, and Tily (2013) and Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017), the optimal random-effects structure justified by the data was identified using forward model selection[3]. The only random slope that significantly improved fit and was included in RT models was that of Block across subjects. Fixed effects were tested using likelihood-ratio tests comparing full and reduced models. All modelling was conducted using the “lme4” package (Bates, Mächler, & Bolker, 2011) in R (R Development Core Team, 2004). Planned contrasts examining the effects of Prime compared balanced/unbalanced homonyms to both sets of non-homonyms. These tests were conducted using the “phia” package (De Rosario-Martinez, 2015), and their significance threshold was adjusted using the Holm-Bonferroni method to further prevent spurious results.

2.2.1 Related trials

In Experiment 1a, there was a significant main effect of Prime in both error rates [$\chi^2(3) = 63.1$, $p < .001$] and RTs [$\chi^2(3) = 39.1$, $p < .001$]. Planned contrasts showed less accurate and slower responses to both balanced (both ps < .001) and unbalanced homonyms (both ps < .001) than to non-homonyms (see Figure 1 below). Responses were generally less accurate and slower to LF-meaning than HF-meaning targets, and this main effect of Target [errors: $\chi^2(1) = 37.9$, $p < .001$; RTs: $\chi^2(1) = 13.7$, $p < .001$] interacted with Prime [errors: $\chi^2(3) = 39.7$, $p < .001$; RTs: $\chi^2(3) = 26.2$, $p < .001$]. Relative to both targets of non-homonyms, responses were less accurate and slower to the LF-meaning targets of balanced (both ps < .001) and unbalanced homonyms (both ps < .001) and the HF-meaning targets of balanced homonyms (errors: p < .01; RTs: p < .05), but not to the HF-meaning targets of unbalanced homonyms (errors: p = .33; RTs: p = .49).

The results of Experiment 1b were very similar. There was a significant main effect of Prime [errors: $\chi^2(3) = 90.5$, $p < .001$; RTs: $\chi^2(3) = 36.9$, $p < .001$]. Planned contrasts showed less accurate and slower responses to both balanced (both ps < .001) and unbalanced homonyms (both ps < .001) than to non-homonyms (see Figure 1 below). Responses were generally less accurate and slower to LF-meaning than HF-meaning targets, and this main effect of Target [errors: $\chi^2(1) = 41.2$, $p < .001$; $\chi^2(1) = 22.0$, $p < .001$] interacted with Prime [errors: $\chi^2(3) = 46.4$, $p < .001$; $\chi^2(3) = 39.0$, $p < .001$]. Relative to both targets of non-homonyms, responses were less accurate and slower to the LF-meaning targets of balanced (both ps < .001) and unbalanced homonyms (both ps < .001) and the HF-meaning targets of balanced homonyms (errors: p < .001; RTs: p < .05), as well as less accurate to the HF-meaning targets of unbalanced homonyms (errors: p < .05; RTs: p = .65).

>> Insert Figure 1 here <<

2.2.2 Unrelated trials

In Experiment 1a, there was only a significant main effect of Prime in error rates [$\chi^2(3) = 10.3$, $p < .05$]. Compared to non-homonyms, responses were less accurate to balanced homonyms but more accurate to unbalanced homonyms, though neither contrast was significant after the Holm-Bonferroni correction (both ps = .10). There was also a significant main effect of Prime in RTs [$\chi^2(3) = 12.8$, $p < .01$]. Compared to non-homonyms, responses were slower to balanced ($p < .01$) but not unbalanced homonyms ($p = .16$; see Figure 2 below).

In Experiment 1b, the analyses revealed an unexpected, marginal Prime × Target interaction in error rates [$\chi^2(3) = 7.7$, $p = .05$] that was due to numerically higher error rates for the LF-meaning targets of balanced homonyms than one of the two sets of targets paired with non-homonyms. This contrast, however, was not significant ($p = .14$) after the Holm-Bonferroni correction. As in Experiment 1a, there was a significant main effect of Prime in RTs [$\chi^2(3) = 16.9$, $p < .001$]. Compared to non-homonyms, responses were slower to balanced ($p < .001$) but not unbalanced homonyms ($p = .10$; see Figure 2 below).

>> Insert Figure 2 here <<

2.3 Discussion

Two key findings emerged from Experiment 1. To begin with, we showed that meaning frequency modulates the ambiguity disadvantage, just like in earlier investigations into sentence reading (Duffy et al., 1988; Rayner & Duffy, 1986). Unbalanced homonymy does not produce the disadvantage, most likely due to weak and delayed activation of the LF meaning in neutral context (Frost & Bentin, 1992; Klepousniotou et al., 2012; Simpson & Burgess, 1985). This is evident in the finding that participants rarely selected the LF meaning, even when there was enough time to do so (high error rates in Experiment 1b), and that a processing cost arose only on trials which instantiated that meaning (higher RTs for unbalanced homonyms than non-homonyms on LF-meaning trials). It appears, then, that unbalanced homonymy slows processing only in rare situations when the dominant meaning turns out to be incorrect, forcing readers to engage in effortful and time-consuming retrieval of the alternative meaning. We revisit this explanation in Experiment 2.

The pattern of responses was different for balanced homonyms. These words incurred a significant processing cost whenever they were encountered (higher RTs for balanced homonyms than non-homonyms on related and unrelated trials). Thus, not only do we confirm that the ambiguity disadvantage in relatedness decision tasks is restricted to balanced homonyms, but we also show, for the first time, that this effect may indeed lie in semantic competition. While the decision-making account (Pexman et al., 2004) assumes ambiguity to slow processing on related but not unrelated trials because only the former involves response conflict, our findings indicate that this is not the case. Experiments 1a and 1b revealed robust ambiguity disadvantage effects for balanced homonyms even on unrelated trials that are free of such conflict. We suspect that meaning frequency may be key to explaining the inconsistencies in findings, especially after discovering that the study by Pexman et al. (2004) included both balanced and unbalanced homonyms within a single stimulus list[4]. It is possible, then, that the study found a null ambiguity effect on unrelated trials because it did not distinguish between the effects of balanced and unbalanced homonymy but combined them instead. Given the present evidence using well-controlled categories of balanced and unbalanced homonymy, the proposal that the ambiguity disadvantage is due to response-selection difficulties on related but not unrelated trials appears to lack support, in that it does not accommodate the findings when the role of meaning frequency in the semantic activation process is taken into account.

Before turning to Experiment 2, it is important to discuss the relatively large proportion of errors on related trials in Experiments 1a and 1b. While errors were expected to be very common for homonyms in the LF-meaning condition (Harpaz, Lavidor, & Goldstein, 2013; Pexman et al., 2004), the results showed that participants made errors even on trials involving non-homonyms. We think that these difficulties in detecting and judging the relatedness between primes and targets were due to multiple constraints in stimulus selection. First, targets were semantic (e.g., “tap-sink”) rather than lexical associates[5] (e.g., “tap-water”), such that participants had to retrieve and consider a number of semantic features of the word referents, which aimed to make the task more sensitive to the impact of semantic activation (Lucas, 2000; Witzel & Forster, 2014). Second, primes and targets were also carefully selected and matched on 14 lexical and semantic properties that have been shown to influence on-line word processing (see Sections 1.1 & 1.2 in the Supplementary Material).

Certain compromises had to be made as a result of these constraints. In particular, we note that some of the primes we used may have not been particularly good at eliciting a given meaning, or as good as they would be when presented together with an associated particle (e.g., “egg” vs. “egg on” in relation to “urge”; “tend” vs. “tend to” in relation to “habit”). We were able to address this issue, however, by demonstrating that high error rates persist and results remain qualitatively similar when these primes are excluded from analyses (see Section 1.4 in the Supplementary Material). Likewise, we note that some of the targets may have been difficult to process in relation to primes because they had multiple semantically related senses themselves (e.g., “tend-nurse” where “nurse” could denote a medical professional, to breast-feed, or to take special care). This was unavoidable given that over 80% of the words in English are ambiguous in this way (Rodd, Gaskell, & Marlsen-Wilson, 2002). We did, however, take this into consideration and controlled for the number of word senses both at the design (Section 1.2 in the Supplementary Material) and the analysis stage (see Sections 2.2 & 2.3 in the Supplementary Material). Thus, although our rigorous control over primes and targets may have contributed to less salient relatedness between the words and less accurate performance[6], we stress that this was instrumental for the design of our study. For example, matching targets for a large number of control variables, rather than letter count and/or word frequency alone (Gottlob et al., 1999; Harpaz et al., 2013; Pexman et al., 2004; Piercey & Joordens, 2000), was necessary to make direct and reliable comparisons of ambiguity effects in different contexts/prime-target combinations.

3 Experiment 2

In Experiment 2, we used EEG measurements to establish when the ambiguity disadvantage for balanced homonyms arises, or at which stage of the relatedness decision performance. Given that our behavioral results lent support to the semantic competition account, we expected to observe the ambiguity disadvantage in the N400 component that has been linked to the ease of semantic processing. In short, the N400 refers to a negatively-going wave that typically peaks

400 ms after the onset of words, pictures, and other meaningful stimuli. Semantic priming, prior context, and predictability have all been shown to attenuate the relative amplitude of the N400 to a word, hence the growing consensus is that the component indexes semantic activation, with larger amplitudes indicating more effortful form-to-meaning mapping (for a review, see Federmeier & Laszlo, 2009). Thus, if balanced homonymy produces competition at the semantic level, as seems to be the case based on our behavioral results, Experiment 2 should show larger N400 amplitudes for balanced homonyms than non-homonyms. It is critical that this effect emerges during the reading of the ambiguous prime, separating early semantic activation processes during prime presentation from late response-selection processes during target presentation.

Experiment 2 also aimed to further examine the processing of unbalanced homonyms. Our behavioral results suggest that these words do not produce semantic competition due to weak and delayed activation of the LF meaning in minimal context. To substantiate this proposal, we compared the amount of priming for the HF and LF meanings of balanced and unbalanced homonyms, focusing again on the N400. The literature on semantic priming has shown that targets preceded by related primes tend to elicit smaller N400 amplitudes than those preceded by unrelated primes (for a review, see Kutas & Federmeier, 2011). This “N400 priming” effect has often been used to investigate patterns of meaning activation in homonyms, both in isolation (e.g., Atchley & Kwasny, 2003; Klepousniotou et al., 2012) and biasing context (e.g., Dholakia, Meade, & Coch, 2016; Swaab, Brown, & Hagoort, 2003). The general finding from such studies is that meanings that are highly frequent or supported by surrounding context are more readily available, and therefore produce greater N400 priming.

Following this literature, we examined N400 effects to related and unrelated targets to determine the extent to which the meanings of homonyms are activated during semantic relatedness decisions. For balanced homonyms, there should be a comparable N400 priming effect for targets instantiating either of the meanings. This would indicate that both meanings are activated to the same extent and in parallel. For unbalanced homonyms, on the other hand, there should be substantial priming for the HF meaning, but little or even no priming for the LF counterpart. This would support our idea that, in isolation or neutral context, readers typically fail to comprehend unbalanced homonyms in the unexpected alternative meaning due to reduced and insufficient activation of that meaning.

3.1 Method

3.1.1 Participants

A different group of 34 University of Leeds students and staff [27 females, aged 18-33 (M = 20.9, SD = 3.5)] participated in exchange for course credit or £8. All participants were right-handed monolingual native British-English speakers with no known history of any language-/vision-related difficulties or neurological damage or disorders. The experiment received ethical approval from the School of Psychology, University of Leeds Ethics Committee.

3.1.2 Stimuli & procedure

Experiment 2 involved the same task and stimuli as Experiment 1b, but there were four minor changes to the procedure. First, participants responded with a computer mouse, rather than a keyboard. Second, there were four, rather than two, one-minute breaks – one before each experimental block. Third, we used a longer inter-trial interval (1000 ms instead of 100 ms) to allow participants to blink and rest their eyes, and there was a 200 ms interval between the target and response execution that aimed to minimize any overlap in ERP components evoked by these trial events. Trials began with a 500 ms fixation cross. After 100 ms, the prime and the target were presented for 700 ms and 500 ms, respectively, with a 50 ms interval in between. Once the target disappeared, there was a 200 ms interval followed by a 1500 ms visual cue ("???") for response execution. Trials ended with a 1000 ms inter-trial-interval (ITI). Fourth, instructions and feedback within the practice block emphasized response accuracy only. Effects in RTs were of no particular interest because Experiment 2 involved a delayed response paradigm, which may have to some extent contaminated our measure of lexical-semantic processing.

3.1.3 EEG data acquisition

The EEG was recorded using 64 pin-type active Ag/AgCl electrodes that were embedded in a head cap, arranged according to the extended 10-20 positioning system (Sharbrough et al., 1991), and connected to a BioSemi ActiveTwo AD-box with an output impedance of less than 1Ω (BioSemi, Amsterdam, the Netherlands). Recording involved 10 midline electrodes and 27 electrodes placed over each hemisphere. Ground electrodes were placed between Cz and CPz. Eye movements were monitored using four electrodes – bipolar horizontal electro-oculogram (EOG) was recorded between the outer right and left canthi, and bipolar vertical EOG was recorded above and below the left eye. Additional electrodes were placed on the left and the right mastoid. The EEG and EOG were recorded continuously with a bandpass filter of 0.16-100 Hz and digitised at a 512-Hz sampling rate.

3.1.4 EEG data pre-processing

The EEG was pre-processed off-line using MATLAB (The Mathworks, Natick, Massachusetts) and EEGLAB (Delorme & Makeig, 2004). The data were first down-sampled to 250 Hz, referenced to the algebraic average of the left and the right mastoid, and then filtered (0.1 - 40 Hz, 12 dB/Oct, Butterworth zero phase filter). Blinks, eye movements, muscle activity, bad channels, and other artifacts were corrected for based on independent component analysis (ICA) guided by measures from SASICA (Chaumon, Bishop, & Busch, 2015; on average, 2-4 components per participant were removed). Cleaned data were then segmented into two types of epochs. For prime-window analyses, epochs started 100 ms before and ended 700 ms after the onset of the prime. For target-window analyses, epochs started 50 ms before the onset of the target and ended 200 ms after the offset. The 100/50 ms intervals before the onset of the prime/target were used to normalize the onset voltage of the ERP waveform.

3.2 Results

3.2.1 Behavioural data

Two of the 34 participants were removed from all analyses – one due to a relatively large number of errors on related (37.1%) and unrelated trials (25.9%) and the other due to a large number of epochs containing amplifier saturation artifacts (+/- 100 µV; 49.0% in the prime window, 54.6% in the target window). Accuracy and latency data were analyzed in the same way as in Experiment 1. For RTs, we excluded errors (13.2% of trials) and any responses that were two SDs above/below a participant's mean in a given condition (4.9% of trials). We did not transform the remaining RTs as the residuals from linear mixed-effects models followed a normal distribution. All models included Prime and Target as fixed effects as well as random intercepts for subjects and items. As in Experiment 1, RT models included Block as an additional fixed effect.

The results were similar to those of Experiments 1a and 1b. For related trials, there was a significant main effect of Prime in both error rates [$\chi^2(3) = 56.7$, $p < .001$] and RTs [$\chi^2(3) = 34.2$, $p < .001$]. Planned contrasts showed less accurate and slower responses to both balanced (both ps < .001) and unbalanced homonyms (both ps < .001) than to non-homonyms (see Figure 3 below). Responses were generally less accurate and slower to LF-meaning than HF-meaning targets, and this main effect of Target [errors: $\chi^2(1) = 32.0$, $p < .001$; RTs: $\chi^2(1) = 23.8$, $p < .001$] interacted with Prime [errors: $\chi^2(3) = 49.3$, $p < .001$; RTs: $\chi^2(3) = 40.5$, $p < .001$]. Relative to both targets of non-homonyms, responses were less accurate and slower to the LF-meaning targets of balanced (both ps < .001) and unbalanced homonyms (both ps < .001), but not to the HF-meaning counterparts of balanced (errors: p = .08; RTs: p = .08) or unbalanced homonyms (errors: p = .56; RTs: p = .44). For unrelated trials, there was only a significant main effect of Prime in RTs [$\chi^2(3) = 24.6$, $p < .001$]. Compared to non-homonyms, responses were slower to balanced ($p < .01$) but not unbalanced homonyms ($p = .10$).

>> Insert Figure 3 here <<

Before turning to EEG data, it is important to note that although Experiments 1 and 2 showed the same patterns of responses for balanced and unbalanced homonyms, the overall error rates appeared to be lower in Experiment 2 (compare Figures 1-3). In order to examine this further, we decided to contrast participants' performance in Experiments 1b and 2 as these were the most similar with respect to stimulus-presentation procedures. The analyses below were the same as those conducted for each experiment separately, except that they included the factor of Experiment (in addition to Prime and Target). All models included significant random intercepts for subjects and items as well as a random slope for Experiment across subjects.

For related trials, there was a significant main effect of Experiment [$\chi^2(1) = 7.5$, $p < .01$], with higher error rates in Experiment 1b ($M = 30.4$, $SD = 8.9$) than Experiment 2 ($M = 24.0$, $SD = 6.2$). There was also a significant Experiment × Prime interaction [$\chi^2(3) = 26.8$, $p < .001$]. Post hoc tests indicated that the simple effect of Experiment (i.e., higher error rates in Experiment 1b) was significant for balanced (Experiment 1b: $M = 30.5$, $SD = 12.4$; Experiment 2: $M = 27.3$, $SD = 8.2$; $p < .001$) and unbalanced homonyms (Experiment 1b: $M = 50.1$, $SD = 11.2$; Experiment 2: $M = 42.7$, $SD = 8.9$; $p < .01$), but not for non-homonyms (Experiment 1b: $M = 15.6$, $SD = 8.1$; Experiment 2: $M = 13.0$, $SD = 6.8$; $p = .35$). The simple effect of Experiment was also significantly greater for balanced than unbalanced homonyms ($p < .001$). For unrelated trials, there was only a significant main effect of Experiment [$\chi^2(1) = 7.7$, $p < .01$], with higher error rates in Experiment 1b (M = 11.9, SD = 2.7) than Experiment 2 (M = 10.4, SD = 2.9).

These results suggest that detecting and judging semantic relatedness was somewhat easier in Experiment 2, especially on trials involving ambiguous words. One particularly important difference between the experiments concerned the instructions given to participants and feedback within the practice block. While in Experiment 1 the instructions and training emphasized both response speed and accuracy, in Experiment 2 they emphasized accuracy only (RTs were of no particular interest due to the delayed response paradigm in the experiment). We think that not only does this explain why accuracy was superior in Experiment 2, but it also sheds some light on our task in general. It appears that the fast-paced nature of our task, or over-emphasis on speed on participants' part, may have to some extent compromised accuracy. This, coupled with the use of more difficult prime-target word pairs, as discussed earlier, could explain why error rates in the present study were relatively high even in the easier conditions involving homonyms in the HF meaning and non-homonyms.

3.2.2 EEG data

EEG analyses excluded individual epochs containing amplifier saturation artifacts (+/- 100 µV; 0.9% of trials in the prime window, 1.3% in the target window) or errors (12.5% of all trials). Following recent studies (Amsel, 2011; De Cat, Klepousniotou, & Baayen, 2015; Kornrumpf, Niefind, Sommer, & Dimigen, 2016), epoched data were analysed on a trial-by-trial basis using linear mixed-effects modelling, primarily due to a large number of errors on LF-meaning trials. As in De Cat et al. (2015), we analyzed each of the 64 channels separately as there was too much data (over 700,000 observations per channel) to fit a single model. In order to prevent spurious results due to a potential multiplicity problem, we used topographical consistency as an additional criterion when judging the reliability of results. The rationale was that since channels are not entirely independent, any effects specific to ambiguity should be similar across neighbouring channels.

Since our hypotheses for both the prime and the target window concerned the N400, analyses focused on the 350-500 ms segment, which best represented this component in our data. Visual inspection of the waveforms within the segment during prime (see Figure 4 below) and target presentation (see Figures 5 & 6 below) revealed a large difference in peak latency for unbalanced homonyms during target presentation (i.e., much earlier peaks for the HF-meaning than LF-meaning/unrelated targets). Thus, as in previous ERP studies of semantic influences on reading (e.g., Taler, Kousaie, & Lopez Zunini, 2013), we divided the 350-500 ms segment into four consecutive time bins of 50 ms in order to capture and account for the divergence in the waveforms.

3.2.2.1 Prime presentation

Prime-window analyses compared N400 amplitudes to homonymous and non-homonymous words during prime presentation. This involved a set of mixed-effects models with the factors of Prime (balanced homonym, unbalanced homonym, non-homonym$_1$, non-homonym$_2$), Time (350-400 ms, 400-450 ms, 450-500 ms, 500-

550 ms), and Block (1, 2, 3, 4). Models included random intercepts for subjects and items as well as random by-subject slopes (mainly for Time). Planned contrasts compared balanced/unbalanced homonyms to both sets of non-homonyms, and their significance level was adjusted using the Holm-Bonferroni method. Only significant effects that involved Prime and were relevant to the hypotheses are reported below.

There was a significant Prime × Time interaction at all channels (all ps < .05), except for T7, TP7, and P9 (for full test results for each channel and effect, see Section 3 in the Supplementary Material). Amplitudes in the 400-450 ms window were larger (i.e., more negative) for balanced homonyms than non-homonyms at fronto-polar (FPz), anterio-frontal (AFz, AF3, AF4, AF8), frontal (Fz, F1, F3, F2, F4), fronto-central (FCz, FC1, FC3, FC2), and fronto-temporal sites (FT7, FT8). This effect also occurred at similar sites in the earlier 350-400 ms (FPz, AF8, Fz, F2, F4, FT7, & FT8; all ps < .05) and the later 450-500 ms window(FPz, AF3, AF4, AF8, Fz, F1, F2, F4, FT7, & FT8; all ps < .05). There were no significant differences between unbalanced homonyms and non-homonyms (see Figure 4 below). Overall, then, the prime-window analyses showed increased negativity from 350 ms to 500 ms post-prime onset for balanced but not unbalanced homonyms. This effect appeared over bilateral medial frontal sites, extending anteriorly to anterio-frontal sites and posteriorly to fronto-temporal sites.

>> Insert Figure 4 here <<

3.2.2.2 Target presentation

Target-window analyses compared N400 amplitudes to the related and unrelated targets of homonyms. This involved a set of mixed-effects models with the factors of Prime (balanced homonym, unbalanced homonym), Target (HF-meaning, LF-meaning, unrelated$_A$, unrelated$_B$), Time (350-400 ms, 400-450 ms, 450-500 ms, 500-550 ms), and Block (1, 2, 3, 4). Non-homonyms were excluded as the aim was to examine the amount of priming for the different meanings of balanced versus unbalanced homonyms. Models included random intercepts for subjects and items and random by-subjects slopes (mainly for Block or Target). Planned contrasts compared HF- and LF-meaning targets to each other and to both sets of unrelated targets, and their significance threshold was adjusted using the Holm-Bonferroni method. Only significant effects that involved Target and were relevant to the hypotheses are reported below.

There was a significant main effect of Target (all ps < .05) at fronto-central (FCz, FC1, FC2), central (Cz, C1, C3, C5, C2, C4), centro-parietal (CPz, CP1, CP3, CP5, CP2, CP4, CP6), parietal (Pz, P1, P3, P5, P2, P4, P6, P8), parieto-occipital (POz, PO3, PO4, PO8), and occipital sites (Oz, O1, O2). Planned contrasts showed reduced (i.e., less negative) amplitudes to HF-meaning targets (all ps < .05) relative to unrelated and LF-meaning targets at most of these channels. There were no significant differences between LF-meaning and unrelated targets.

There was a significant Target × Time interaction (all ps < .05) at all channels, except for P9. HF-meaning targets elicited smaller amplitudes than unrelated targets in the 400-450 ms, 450-500 ms, and 500-550 ms windows at all the channels (all ps < .05), except for AF7, AF8, F5, F7, F8, FT7, T7, TP7, P7, and P10. In addition, HF-meaning targets elicited smaller amplitudes than LF-meanings targets in the 500-550 ms window at frontal (Fz, F2), fronto-central (FCz, FC1, FC2, FC4), central (Cz, C1, C3, C5, C2, C4), centro-parietal (CP1, CP3, CP5, CP2, CP4, CP6), parietal (Pz, P1, P3, P5, P2, P4, P6), parieto-occipital (POz, PO3, PO4), and occipital sites (Oz, O1, O2), as well as the inion (Iz; all ps < .05). This effect was also significant in the earlier 400-450 ms and 450-500 ms windows at the same channels (all ps < .05), except for F2, FC4, C1, C5, C4, CP3, CP5, CP6, P2, and Iz. LF-meaning targets, on the other hand, elicited smaller amplitudes than unrelated targets at CP4, CP6, P2, POz, and PO8 in the 450-500 ms window only (all ps < .05).

There was a significant Target × Time × Prime interaction (all ps < .05) at all channels, except for T7, TP7, P7, and P9. For balanced homonyms (see Figure 5 below), HF-meaning targets elicited smaller amplitudes than unrelated targets in the last 500-550 ms window at fronto-central (FCz, FC1, FC3, FC2, FC4), central (Cz, C1, C3, C5, C2, C4), centro-parietal (CPz, CP1, CP3, CP5, CP2, CP4, CP6), parietal (Pz, P1, P3, P5, P2, P4, P6, P8), and parieto-occipital sites (POz, PO3, PO4; all ps < .05). This effect also occurred in the earlier 450-500 ms window at a smaller cluster (Cz, CPz, CP1, CP2, CP6, Pz, P1, P2, POz, PO3, & PO4; all ps < .05). The contrasts between the HF-meaning and LF-meaning targets as well as between the LF-meaning and unrelated targets for balanced homonyms were not significant.

>> Insert Figure 5 here <<

For unbalanced homonyms (see Figure 6 below), on the other hand, HF-meaning targets elicited smaller amplitudes than unrelated targets in the 450-500 ms and 500-550 ms windows at all the channels (all ps < .05), except for AF7, AF8, F5, F7, FT7, and P10 (in addition to the four channels that did not show the 3-way interaction in the first place). This effect also occurred in the earlier 400-450 ms window (all ps < .05) at the same channels, except for AF3, FP1, F3, FC5, TP8, CP5, P8, PO7, and Iz. In addition, HF-meaning targets elicited smaller amplitudes than LF-meaning targets in the last 500-550 ms window at fronto-polar (FPz, FP1, FP2), anterio-frontal (AFz, AF3, AF4), frontal (Fz, F1, F2, F4, F6), fronto-central (FCz, FC1, FC3, FC2, FC4), central (Cz, C1, C3, C2, C4, C6), centro-parietal (CPz, CP1, CP2), parietal (Pz, P1, P3, P2, P4), parieto-occipital (POz, PO3, PO4), and occipital sites (Oz, O1, & O2; all ps < .05). This effect also occurred at similar sites in the earlier 400-450 ms (FPz, Fz, F4, F6, FCz, FC1, FC2, Cz, C1, C2, C6, CPz, CP1, CP2, Pz, P1, P3, PO3, Oz, & O2) and 450-500 ms windows (AFz, AF3, AF4, FPz, FP1, FP2, Fz, F1, F4, F6, FCz, FC1, FC3, FC2, FC4, Cz, C1, C2, C4, C6, CPz, CP1, CP2, Pz, P1, P3, PO3, Oz, & O2; all ps < .05). The contrasts between the LF meanings and unrelated targets for unbalanced homonyms were not significant.

In summary, the target-window analyses showed that amplitudes to the HF-meaning targets of balanced homonyms were reduced only in comparison to unrelated targets, primarily from 500 ms to 550 ms post-target onset. Amplitudes to the HF-meaning targets of the unbalanced counterparts, on the other hand, were reduced in comparison to both unrelated and LF-meaning targets, and this effect was markedly sustained (400-550 ms post-target onset). For both balanced and unbalanced homonyms, priming for the HF meaning appeared over bilateral medial and lateral centro-parietal sites, extending anteriorly to frontal sites and posteriorly to occipital sites. In contrast, no significant differences were observed between LF-meaning and unrelated targets for either balanced or unbalanced homonyms.

>> Insert Figure 6 here <<

3.3 Discussion

Two key findings emerged from Experiment 2. To begin with, analyses for the prime window revealed increased frontal negativity from 350 ms to 500 ms post-prime-onset for balanced but not unbalanced homonyms (relative to non-homonyms), which, as we argue in the section below, is compatible with the semantic competition account (Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004). The finding that homonymy in general had an impact on the N400 is consistent with previous lexical decision studies which reported greater N400 responses to homonyms than non-homonyms (Haro, Demestre, Boada, & Ferré, 2017; see also Beretta, Fiorentino, & Poeppel, 2005; MacGregor et al., 2020). Not only does our experiment corroborate and extend this work by demonstrating that homonymy also affects the N400 component in semantically engaging tasks that require disambiguation, but it also shows that it is balanced, not unbalanced, homonymy that drives this effect. In other words, our study is the first to provide EEG evidence for the long-held assumption that meaning frequency modulates ambiguity effects in word processing (for behavioral evidence, see Experiment 1; Armstrong et al., 2012; Brocher et al., 2018; Rayner & Duffy, 1986).

Turning to the analyses for the target window, the results confirmed that balanced and unbalanced homonyms differ in the extent to which their meanings are activated in the absence of context. There was a significant N400 priming effect for HF-meaning targets and a non-significant one for LF-meaning targets (relative to unrelated targets), both for balanced and unbalanced homonyms. Note, however, that there was evidence to suggest that (weaker) priming also occurred for the LF meaning of balanced homonyms. Targets instantiating that meaning elicited N400 amplitudes that were (a) numerically, though not statistically, smaller than those for unrelated targets and (b) comparable to those for HF-meaning targets (see Figure 5 above). In other words, while the dominant meaning was activated and facilitated the processing of the related target for both types of homonyms, the alternative counterpart was activated (to a lesser degree) only for balanced homonyms. This suggests that balanced and unbalanced homonyms differ in how and when their meanings are activated, and may therefore produce different levels of semantic competition.

4 General Discussion

The present study provides consistent behavioral and electrophysiological evidence that the ambiguity disadvantage is due to competition between multiple semantic representations during the activation process, as predicted by current connectionist models of ambiguity (Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004). Experiment 1 shows that the ambiguity disadvantage arises for balanced but not unbalanced homonyms, extending previous findings from sentence reading (Duffy et al., 1988; Rayner & Duffy, 1986) to single-word processing. This effect was not restricted to related trials, which proves particularly challenging for the decision-making account proposed by Pexman et al. (2004). While the account assumes ambiguity to slow relatedness decisions on related but not unrelated trials because only the former involves response conflict, we demonstrate that this is not the case - balanced homonymy incurred a processing cost regardless of whether the different meanings triggered consistent or inconsistent responses to the target (i.e., both on unrelated and related trials). This is in line with our recent finding that learning new meanings for familiar words slows the processing of their existing meanings (mirroring the ambiguity disadvantage in natural language), both on related and unrelated trials in relatedness decision tasks (Maciejewski, Rodd, Mon-Williams, & Klepousniotou, 2019). Overall, then, it appears that the idea that the ambiguity disadvantage is due to additional decision making involved in response-conflict resolution faces a major challenge, in that it cannot explain why balanced homonymy would incur a processing cost on unrelated trials that are free of response conflict.

Further evidence against the decision-making account comes from Experiment 2. To begin with, the finding that the effect of balanced homonymy was observed during prime presentation confirms that the ambiguity disadvantage arises when processing the ambiguous word itself, rather than when processing or responding to the target. More specifically, it arises during the semantic activation process, as revealed by increased negativity in the N400 window. Note, however, that while the latency of our effect is consistent with that of a typical N400 effect, this is not the case with respect to scalp topography. The ERP literature (for a review, see Kutas & Federmeier, 2011) shows that the “traditional” N400 effect is normally largest over centro-parietal sites, rather than frontal sites as in the current study, though there have been reports of increased frontal negativity for homonyms versus non-homonyms before (Lee & Federmeier, 2006, 2009; see also Mollo, Jefferies, Cornelissen, & Gennari, 2018). This striking difference in topography suggests that the common explanation for an N400 effect in terms of differences in the extent of semantic activation or priming may not fully apply to our effect. Increased frontal negativity for balanced homonymy may instead point to an additional, inhibitory process involved in ambiguity resolution – most likely semantic competition, as suggested by the literature reviewed next.

fMRI studies of ambiguity found that the left inferior frontal gyrus (LIFG), in particular pars triangularis (BA 45) and pars opercularis (BA 44), is the most consistent brain region to show an increased haemodynamic response to ambiguity (for a detailed review, see Vitello & Rodd, 2015), though there is also evidence for bilateral recruitment of that area when processing ambiguous words (Klepousniotou, Gracco & Pike, 2014). There also appears to be wide agreement in this literature that the LIFG is involved in the resolution of competition between the multiple meanings of an ambiguous word, either when the word is encountered in isolation (e.g., Bilenko, Grindrod, Myers, & Blumstein, 2009; Hargreaves et al., 2011) or when the word must be reinterpreted following initial selection of the incorrect meaning (e.g., Rodd, Johnsrude, & Davis, 2012). The former situation closely corresponds to the prime window in Experiment 2, hence increased frontal negativity for balanced homonyms in that window may indicate competition between their meanings within the LIFG. This interpretation is further supported by the influential “conflict resolution” account of LIFG function (e.g., Novick, Trueswell, & Thompson-Schill, 2009; Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997), according to which posterior LIFG serves to resolve competition between multiple representations. In particular, Novick et al. (2009) proposed that posterior LIFG engages in the resolution of competition, regardless of its specific linguistic form, either when there is a prepotent but irrelevant response, or when there are multiple activated representations but no dominant response. Since reading balanced homonyms in Experiment 2 produced the latter type of competition, at the semantic level in this case, increased frontal negativity for these words may indeed reflect increased activation of the LIFG, and its RH homologue, in an attempt to resolve that competition[7].

Overall, then, the present findings are incompatible with the idea that response conflict constitutes an explanation for the ambiguity disadvantage (Pexman et al., 2004). In particular, the finding that balanced homonymy affected the N400 component in the prime window indicates that the effect arises during the semantic activation of the ambiguous prime, hundreds of milliseconds before participants see the related/unrelated target that follows. This clearly shows that the ambiguity disadvantage is not due to response-selection difficulties upon target presentation, but due to semantic competition in response to ambiguity itself. Note, however, that the present findings are not necessarily incompatible with Hino et al.'s (2006) decision-making account that focuses on qualitative task differences and their impact on how the response system is configured, rather than response conflict. Under this account, the ambiguity disadvantage arises only when a task places demands on post-semantic processes, such as analysis of semantic features, that support response selection. We do not provide compelling evidence either for or against this account, since our study aimed to examine the ambiguity disadvantage in semantic relatedness decisions, rather than semantic categorisations that the account focuses on. We do, however, think that it is possible to marry some aspects of the decision-making account with the semantic competition one. In particular, we agree with Hino et al. (2006) that the impact of ambiguity in word processing largely depends on what readers/listeners must do with the word. For instance, in relatedness decision tasks, competition effects arise for homonyms because responses are made based on complete semantic activation, in the sense that participants must settle on a particular meaning of these words. In lexical decision tasks, competition effects become noticeable only when responses are more reliant on semantic activation, when, for example, discriminating between homonyms and pseudo-homophonic (Azuma & Van Orden, 1997; Rodd et al., 2002) or wordlike non-words (Armstrong & Plaut, 2008, 2016). Likewise, in semantic categorization tasks, competition effects arise only when there is a need for greater semantic activation, when responses cannot be made based on a small number of semantic features (Hino et al., 2006). Thus, the general idea is that task demands play some role in generating ambiguity effects. However, while Hino et al. (2006) assert that this is due to differences in how the response system is configured in a particular task, we suggest that this is more likely due to differences in the level of semantic activation needed to perform the task (for a similar view, see Armstrong & Plaut, 2016). This is supported by our demonstrations that meaning frequency, which influences the level of semantic activation, modulates the ambiguity disadvantage, and that the ambiguity disadvantage arises during semantic, rather than post-semantic, processes.

In contrast, the present findings are readily compatible with the predictions of current connectionist models of ambiguity (Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004). Explaining why meaning frequency would modulate the ambiguity disadvantage presents these models with little challenge. Within the connectionist framework, long-term experience with a particular meaning of an ambiguous word modifies the strength of the connections between orthographic and semantic units, which in turn determines the speed and outcome of form-to-meaning mapping. The HF meanings of unbalanced homonyms develop strong connections, and thus are activated so fast that they avoid competition with the LF meanings. For balanced homonyms, meaning frequency plays barely any role in form-to-meaning mapping - both meanings are activated to the same extent and in parallel, with each being equally likely to win competition for further activation. Therefore, connectionist models of ambiguity, such as the one implemented by Kawamoto (1993), can easily account for the differential effects of balanced and unbalanced homonymy in semantic activation (and competition involved) by modifying the weights on the connections between orthographic and semantic units.

Note that these differential effects are evident in both our behavioral and electrophysiological data. For unbalanced homonyms, activation of the LF meaning was so weak that participants rarely selected that meaning in minimal context, even when there was enough time do so. This is in line with the finding of very high error rates for unbalanced homonyms in the LF meaning in the short (Experiment 1a) and the long prime-duration condition (Experiments 1b & 2). When participants did disambiguate the words towards the LF meaning, there was a substantial processing cost (higher RTs for unbalanced homonyms than non-homonyms on correct LF-meaning trials in Experiments 1 & 2) that we take as evidence of effortful and slow retrieval of that meaning upon seeing a supporting target. This is in line with the finding of no N400 priming for the LF-meaning target even on correct trials (Experiment 2) as well as the finding of a similar processing cost at unexpected LF-meaning context following an unbalanced homonym (e.g., “We knew the boxer was barking all night”) in eye-movement studies (Duffy et al., 1988; Rayner & Duffy, 1986).

Importantly, the difficulty in processing the LF meaning of unbalanced homonyms did not arise because participants did not know that meaning. Maciejewski and Klepousniotou’s (2016) norms, from which the homonyms were selected, confirm that over 75 out of the 100 native speakers they tested used and/or encountered the LF meaning of these words[8]. This suggests that, for most participants in the current study, the meaning was stored in the mental lexicon but not sufficiently activated in the absence of context. Support for this interpretation comes from the finding that readers struggle but eventually manage to understand unbalanced homonyms in the LF meaning solely based on strong sentential context (e.g., Brocher et al., 2018; Duffy et al., 1988; Leinenger, Myslín, Rayner, & Levy, 2017). Further support comes from stimulus pre-tests that we conducted as part of the study (see Section 1.3 in the Supplementary Material). These pre-tests showed that raters normally failed to detect the semantic relatedness between unbalanced homonyms and targets instantiating the LF meaning, unless they were first presented with sentential context supporting that meaning. The implication is that naturalistic and elaborate context may be necessary to fully retrieve and select the LF counterpart, both in on-line and off-line tasks. This is because, for ease of comprehension, the language system appears to process unbalanced homonyms as functionally unambiguous words.

For balanced homonyms, the results indicate that although both their meanings were sufficiently activated to produce semantic competition, they did not seem to be activated to the same extent. After all, there were fewer errors

(Experiments 1 & 2), faster responses (Experiments 1 & 2), and larger N400 priming (Experiment 2) on HF-meaning than LF-meaning trials. This should not come as a surprise. Truly balanced homonyms are very rare at best (see Armstrong et al., 2012), hence the relative frequencies of the meanings of our words differed, on average, by 20% (SD = 12). It appears, then, that even balanced homonyms show small, albeit noticeable, bias in the activation process.

Lastly, we wish to emphasize that although our study lends support to the semantic competition account, it does not really help to distinguish between specific connectionist models of ambiguity that proposed the account (Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004). While all three models predict homonymy to produce semantic competition in tasks that require meaning selection (e.g., semantic relatedness decisions), they disagree quite substantially on the impact of homonymy in tasks that do not (e.g., lexical decisions). Kawamoto's (1993) model predicts a facilitatory effect due to enhanced feedback from semantics during orthographic processing, which is at odd with most lexical decision studies (for a review, see Eddington & Tokowicz, 2015). Rodd et al.'s (2004) model, on the other hand, predicts an inhibitory effect due to inconsistent form-to-meaning mappings during semantic processing, regardless of task demands. Armstrong and Plaut's (2008) model also predicts an inhibitory effect, but only when the task is sufficiently difficult to engage substantial semantic processing. Not only do the models disagree on why and when ambiguity has its effect, but they also differ in terms of descriptions of the roles of context, meaning frequency, and meaning relatedness. Kawamoto's (1993) model simulates the predicted effects of meaning frequency but does not make the important distinction between homonyms and polysemes (i.e., words with multiple related senses, such as “nurse”). Rodd et al.’s (2004) and Armstrong and Plaut’s (2008) models make the distinction, but only the latter discusses (but does not simulate) the roles of context and meaning frequency. Taken together, while our study supports the overall semantic competition account, more evidence is needed to advance or constrain the models. In particular, future studies should attempt to extend our findings to other forms of ambiguity, given growing evidence that for polysemes semantic competition may largely depend on the degree of overlap of the multiple senses (Windisch Brown, 2008; Klepousniotou, Titone, & Romero, 2008; Maciejewski et al., 2019), such that it could be minimal for words with highly overlapping senses (e.g., “dust”) but strong, albeit not as much as for homonyms, for words with less overlapping senses (e.g., “virus”).In conclusion, the present findings demonstrate that the ambiguity disadvantage in relatedness decision tasks is restricted to balanced homonyms and show, for the first time, that this effect arises during the semantic processing of the ambiguous word itself. More specifically, the study suggests that balanced homonyms give rise to competition during the semantic activation process which most likely engages the LIFG that has been implicated in the resolution of such competition (for a review, see Vitello & Rodd, 2015). In addition, the study provides direct evidence that balanced and unbalanced homonyms differ in how their meanings are activated out of context, which determines the degree of competition they produce.

The present findings are consistent with semantic competition accounts proposed by connectionist models of ambiguity, especially those that incorporate an explanation for the role of meaning frequency (Kawamoto, 1993). They are not, however, consistent with decision-making accounts, especially those that attribute the ambiguity disadvantage to response-conflict resolution (Pexman et al., 2004). In particular, such accounts fail to accommodate the finding of the ambiguity disadvantage during the semantic processing of the ambiguous word itself, rather than during the processing of the related/unrelated target and subsequent response making. Furthermore, if the ambiguity disadvantage is merely a task artifact at the response-selection stage, it remains unclear why it would be robust across a number of tasks of distinct response-selection demands. After all, competition effects in ambiguity resolution have been observed in tasks involving semantic relatedness (e.g., Gottlob et al., 1999) and categorisation decisions (e.g., Jager & Cleland, 2015), semantically primed (e.g., Klepousniotou, 2002) and unprimed lexical decisions (e.g., Armstrong & Plaut, 2016), sensicality judgements (e.g., Klepousniotou et al., 2008), and even sentence-reading tasks that do not require any response or decision (e.g., Duffy et al., 1988). Our study marks a significant step towards unravelling the locus of these competition effects, in that it establishes that, at least in relatedness decision tasks, these effects arise due to semantic activation processes.

Acknowledgements

This work was funded by an Economic and Social Research Council doctoral studentship (ES/J500215/1) and a University of Leeds postgraduate research grant awarded to the first author. We thank Rebecca Gardner, Marketa Provodova, and Timothy Carling for their help with data collection, and Brian Scully for his help with data pre-processing. We also thank three anonymous reviewers for their helpful comments on an earlier draft of the manuscript.

References

Amsel, B. D. (2011). Tracking real-time activation of conceptual knowledge using single-trial event-related potentials. Neuropsychologia, 49(5), 970-983.

Annett, M. (1967). The binomial distribution of right, mixed and left handedness. The Quarterly Journal of Experimental Psychology, 19(4), 327-333.

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 273-278). Austin, TX: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. Language, Cognition, and Neuroscience, 1(7), 1-27.

Armstrong, B. C., Tokowicz, N., & Plaut, D. C. (2012). eDom: Norming software and relative meaning frequencies for 544 English homonyms. Behavior Research Methods, 44(4), 1015-1027.

Atchley, R. A., & Kwasny, K. M. (2003). Using event-related potentials to examine hemispheric differences in semantic processing. Brain and Cognition, 53(2), 133-138.

Azuma, T., & Van Orden, G. C. (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. Journal of Memory and Language, 36(4), 484-504.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255-278.

Bates, D., Mächler, M., & Bolker, B., (2011). Lme4: Linear mixed-effects modeling using S4 classes [R package]. Retrieved from http://CRAN.R-Project.org/package=lme4.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. Cognitive Brain Research, 24(1), 57-65.

Bilenko, N. Y., Grindrod, C. M., Myers, E. B., & Blumstein, S. E. (2009). Neural correlates of semantic competition during processing of ambiguous words. Journal of Cognitive Neuroscience, 21(5), 960-975.

Briggs, G. G., & Nebes, R. D. (1975). Patterns of hand preference in a student population. Cortex, 11(3), 230-238.

Brocher, A., Koenig, J. P., Mauner, G., & Foraker, S. (2018). About sharing and commitment: The retrieval of biased and balanced irregular polysemes. Language, Cognition and Neuroscience, 33(4), 443-466.

Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. Journal of Neuroscience Methods, 250, 47-63.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9-21.

De Cat, C., Klepousniotou, E., & Baayen, R. H. (2015). Representational deficit or processing effect? An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English. Frontiers in Psychology, 6, 77.

De Rosario-Martínez, H. (2015). Phia: Post-hoc interaction analysis [R package]. Retrieved from http://cran.r-project.org/web/packages/phia.

Dholakia, A., Meade, G., & Coch, D. (2016). The N400 elicited by homonyms in puns: Two primes are not better than one. Psychophysiology, 53(12), 1799-1810.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. Journal of Memory and Language, 27(4), 429-446.

Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: The current state of the literature. Psychonomic Bulletin & Review, 22(1), 13-37.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. In B. H. Ross (Ed.), Psychology of learning and motivation (pp. 1-44). San Diego, CA: Elsevier.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. Journal of Memory and Language, 29(2), 181-200.

Frost, R., & Bentin, S. (1992). Processing phonological and semantic ambiguity: Evidence from semantic priming at different SOAs. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(1), 58-68.

Grindrod, C. M., Garnett, E. O., Malyutina, S., & den Ouden, D. B. (2014). Effects of representational distance between meanings on the neural correlates of semantic ambiguity. Brain and Language, 139, 23-35.

Gottlob, L. R., Goldinger, S. D., Stone, G. O., & Van Orden, G. C. (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. Journal of Experimental Psychology: Human Perception and Performance, 25(2), 561-574.

Hargreaves, I. S., Pexman, P. M., Pittman, D. J., & Goodyear, B. G. (2011). Tolerating ambiguity: Ambiguous words recruit the left inferior frontal gyrus in absence of a behavioral effect. Experimental Psychology, 58(1), 19-30.

Haro, J., Demestre, J., Boada, R., & Ferré, P. (2017). ERP and behavioral effects of semantic ambiguity in a lexical decision task. Journal of Neurolinguistics, 44, 190-202.

Harpaz, Y., Lavidor, M., & Goldstein, A. (2013). Right semantic modulation of early MEG components during ambiguity resolution. Neuroimage, 82, 107-114.

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? Journal of Memory and Language, 55(2), 247-273.

Hoffmann, S., & Evert, S. (2006). BNCweb (CQP-edition): The marriage of two corpus tools. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods (pp. 177-195), Frankfurt: Peter Lang.

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. Journal of Experimental Psychology: Human Perception and Performance, 41(2), 385-402.

Jager, B., & Cleland, A. A. (2015). Connecting the research fields of lexical ambiguity and figures of speech: Polysemy effects for conventional metaphors and metonyms. The Mental Lexicon, 10(1), 133-151.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. Journal of Memory and Language, 32(4), 474-516.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. Brain and Language, 81(1), 205-223.

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. Journal of Neurolinguistics, 20(1), 1-24.

Klepousniotou E., Gracco V.L., & Pike G.B. (2014). Pathways to lexical ambiguity: fMRI evidence for bilateral fronto-parietal involvement in language processing. Brain and Language, 131, 56-64.

Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. Brain and Language, 123(1), 11-21.

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(6), 1534-1543.

Kornrumpf, B., Niefind, F., Sommer, W., & Dimigen, O. (2016). Neural correlates of word recognition: A systematic comparison of natural reading and rapid serial visual presentation. Journal of Cognitive Neuroscience, 28(9), 1374-1391.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review of Psychology, 62, 621-647.

Lee, C. L., & Federmeier, K. D. (2006). To mind the mind: An event-related potential study of word class and semantic ambiguity. Brain Research, 1081(1), 191-202.

Lee, C. L., & Federmeier, K. D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. Journal of Memory and Language, 61(4), 538-555.

Leinenger, M., Myslín, M., Rayner, K., & Levy, R. (2017). Do resource constraints affect lexical processing? Evidence from eye movements. Journal of Memory and Language, 93, 82-103.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. Psychonomic Bulletin & Review, 1(4), 476-490.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. Psychonomic Bulletin & Review, 7(4), 618-630.

MacGregor, L. J., Bouwsema, J., & Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. Neuropsychologia, 68,126-138.

MacGregor, L. J., Rodd, J. M., Gilbert, R. A., Hauk, O., Soboglu, E., & Davis, M. H. (2020). The neural time course of semantic ambiguity resolution in speech comprehension. Journal of Cognitive Neuroscience, 32(3), 403-425.

Maciejewski, G., & Klepousniotou, E. (2016). Relative meaning frequencies for 100 homonyms: British eDom norms. Journal of Open Psychology Data, 4, e6.

Maciejewski, G., Rodd, J. M., Mon-Williams, M., & Klepousniotou, E. (2019). The cost of learning new meanings for familiar words. Language, Cognition and Neuroscience, 1-23.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. Journal of Memory and Language, 94, 305-315.

Mirman, D., Strauss, T. J., Dixon, J. A., & Magnuson, J. S. (2010). Effect of representational distance between meanings on recognition of ambiguous spoken words. Cognitive Science, 34(1), 161-173.

Mollo, G., Jefferies, E., Cornelissen, P., & Gennari, S. P. (2018). Context-dependent lexical ambiguity resolution: MEG evidence for the time-course of activity in left inferior frontal gyrus and posterior middle temporal gyrus. Brain and Language, 177, 23-36.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. Behavior Research Methods, Instruments, & Computers, 36(3), 402-407.

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. Cognitive Neuropsychology, 26(6), 527-567.

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(6), 1252-1270.

Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. Memory & Cognition, 28(4), 657-666.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(3), 785-794.

R Development Core Team (2004). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition, 14(3), 191-201.

Rice, C. A., Beekhuizen, B., Dubrovsky, V., Stevenson, S., & Armstrong, B. C. (2019). A comparison of homonym meaning frequency estimates derived from movie and television subtitles, free association, and explicit ratings. Behavior Research Methods, 51(3), 1399-1425.

Rodd, J. M. (2018). Lexical ambiguity. In S.-A. Rueschemeyer, & M. G. Gaskell (Eds.) Oxford handbook of psycholinguistics (2nd Edition, pp. 120-144). Oxford, UK: Oxford University Press.

Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. Journal of Memory and Language, 46(2), 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(1), 89-104.

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. Cerebral Cortex, 22(8), 1761-1773.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). E-Prime 2.0. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Seidenberg, M. S. (2007). Connectionist models of reading. In M. G. Gaskell (Ed.), Oxford handbook of psycholinguistics (pp. 235-250). Oxford, UK: Oxford University Press.

Sharbrough, F., Chatrian, G.-E., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. Journal of Clinical Neurophysiology, 8, 200-202.

Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. Brain and Language, 86(2), 326-343.

Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. Journal of Experimental Psychology: Human Perception and Performance, 11(1), 28-39.

Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. Journal of Memory and Language, 30(6), 627-643.

Taler, V., Kousaie, S., & Lopez Zunini, R. (2013). ERP measures of semantic richness: The case of multiple senses. Frontiers in Human Neuroscience, 7, 5.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A re-evaluation. Proceedings of the National Academy of Sciences, 94(26), 14792-14797.

Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. Psychonomic Bulletin & Review, 7(1), 49-82.

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. Memory & Cognition, 22(1), 111-126.

Vitello, S., & Rodd, J. M. (2015). Resolving semantic ambiguities in sentences: Cognitive processes and brain mechanisms. Language and Linguistics Compass, 9(10), 391-405.

Windisch Brown, S. (2008). Polysemy in the mental lexicon. Colorado Research in Linguistics, 21(1), 1-12.

Witzel, J., & Forster, K. (2014). Lexical co-occurrence and ambiguity resolution. Language, Cognition and Neuroscience, 29(2), 158-185.

Tables

Table 1. Examples of prime-target word pairs used in Experiments 1 and 2.

Prime	Related target		Unrelated target	
	HF-meaning	LF-meaning	A	B
Balanced homonym	fan-cheer	fan-breeze	fan-snake	fan-cancel
Unbalanced homonym	pen-ink	pen-farmer	pen-yeast	pen-add
Non-homonym$_1$	fake-truth	fake-fraud	fake-expand	fake-fetch
Non-homonym$_2$	fur-fox	fur-rabbit	fur-chain	fur-pill

Footnotes

[1] RT analyses involving log transformation but not SD-based trimming produced qualitatively similar results.

[2] We report the results for Block and discuss why our experiments did not show practice/prime-repetition effects in Section 2.1 in the Supplementary Material.

[3] We began analysis with a model that included random intercepts and tested all possible slopes for inclusion separately. Out of significant slopes, we first added the most influential one (based on the value of χ^2 from model-comparison tests) to the base model and then tested whether the second most influential slope further improved the model. We continued to test and include slopes until the model failed to converge.

[4] To determine the number of balanced and unbalanced homonyms in Pexman et al.'s study (2004, Experiments 1-4), we used Twilley, Dixon, Taylor, and Clark's (1994) meaning-frequency ratings in Canadian English - the dialect spoken by the recruited participants. We found that half of the homonyms had a highly dominant meaning (i.e., meaning frequencies for these words differed by 41-79%), which supports our claim that the study used both balanced and unbalanced homonyms but did not distinguish between them. Note, however, that these are estimates only, in that there may be slight differences in meaning-frequency ratings depending on whether they are derived from television subtitles, word associations, or explicit judgements (see Rice, Beekhuizen, Dubrovsky, Stevenson, & Armstrong, 2019).

[5] We used BNCweb (CQP-edition; Hoffmann & Evert, 2006) to examine how often primes and targets co-occurred within spoken and written language, up to four words apart. This analysis confirmed that all but three related targets were rarely used together with primes in natural discourse, and that our stimuli were not lexical associates.

[6] On average, only two of the 28 word pairs in each condition were forward- (e.g., “tent” in response to “camp”) or backward-generated associates (e.g., “camp” in response to “tent”) in the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 2004). This indicates that primes and targets, regardless of the condition, did not elicit each other’s meanings in a typical, straightforward way.

[7] Note that the scalp topography of ERPs does not allow us to make definitive claims about the localization of neural sources.

[8] The results for unbalanced homonyms were qualitatively similar after removing some of the unbalanced homonyms with lesser-known LF meanings (see Section 2.4 in the Supplementary Material).

Figure Captions

Figure 1. Subject means of % error rates and untransformed RTs in ms for related trials in Experiments 1a (Panel A) and 1b (Panel B). Error rates show 95% confidence intervals adjusted to remove between-subjects variance (Loftus & Masson, 1994).

Figure 2. Subject means of % error rates and untransformed RTs in ms for unrelated trials in Experiments 1a (Panel A) and 1b (Panel B). Error rates show 95% confidence intervals adjusted to remove between-subjects variance.

Figure 3. Subject means of % error rates and untransformed RTs in ms for related (Panel A) and unrelated trials (Panel B) in Experiment 2. Error rates show 95% confidence intervals adjusted to remove between-subjects variance.

Figure 4. Grand average waveforms for balanced/unbalanced homonyms and non-homonyms during prime presentation (at major frontal, central, & posterior locations). Negative amplitudes are plotted downwards.

Figure 5. Grand average waveforms for the HF-meaning, LF-meaning, and unrelated targets of balanced homonyms during target presentation (at major frontal, central, & posterior locations). Negative amplitudes are plotted downwards.

Figure 6. Grand average waveforms for the HF-meaning, LF-meaning, and unrelated targets of unbalanced homonyms during target presentation (at major frontal, central, & posterior locations). Negative amplitudes are plotted downwards.

Appendix

Sets of prime-target words pairs used in Experiments 1 and 2.

Prime		Related pairs		Unrelated pairs	
		HF target	LF target	Target A	Target B
Balanced homonym	bay	creek	alcove	tune	ride
	bust	breast	burst	basil	eat
	calf	knee	cattle	trench	bitter
	camp	tent	gay	lag	quick
	fan	cheer	breeze	snake	cancel
	forge	advance	hammer	bird	pig
	jam	knife	tight	oval	devil
	lean	bend	slim	crime	roar
	novel	poem	unique	wipe	reward
	palm	wrist	exotic	sing	mile
	pine	oak	desire	cloak	stroll
	plot	writer	acre	curl	plug
	prop	pillar	actor	parrot	dinner
	pupil	lesson	lens	enter	pan
	rank	fifth	odour	device	rift
	scrap	pieces	argue	castle	beach
	seal	swim	glue	rapid	monk
	shed	hut	skin	fight	dance
	squash	sports	potato	alive	anchor
	stall	delay	sell	lip	veil
	strip	naked	ribbon	pond	eagle
	tap	sink	knock	beans	poet
	temple	chapel	brow	swan	album
	tend	habit	nurse	begin	insect
	tense	stress	grammar	cook	tea
	toast	dish	beer	skull	ache
	utter	aloud	absolute	fence	sister
	yard	grass	inch	invite	betray
Unbalanced homonym	angle	maths	fisher	bronze	laugh
	cape	jacket	ocean	error	mental

	chord	song	circle	zoo	sore
	corn	crop	toe	preach	quit
	ear	listen	cereal	shelf	excess
	egg	goose	urge	boot	ankle
	fleet	navy	swift	smart	ale
	flock	herd	fabric	screen	skill
	fry	butter	infant	clay	sign
	hide	buried	animal	cheap	acid
	host	guest	plenty	sand	throat
	lock	shut	comb	pest	saint
	mate	pal	chess	galaxy	crust
	mint	ginger	coin	chin	mess
	pad	cloth	foot	anger	frozen
	pen	ink	farmer	yeast	add
	pit	dig	cherry	gaze	sting
	pool	bath	resource	tongue	blade
	pulse	vein	seed	milk	gender
	pump	flow	shoes	hunt	jaw
	rail	barrier	protest	willow	foam
	ray	shine	fish	ripe	coal
	sheer	thin	veer	fridge	nose
	spray	mist	flower	rival	pigeon
	stern	strict	boat	gift	bin
	toll	levy	bell	focus	mud
	verse	poetry	tutor	wet	jungle
	wax	warm	moon	dog	heaven
Non-homonym Set 1	bald	hairy	wig	vocal	ton
	bulk	huge	vast	wait	funny
	crew	squad	crowd	arrow	snow
	curve	chart	graph	guard	flood
	drain	dry	liquid	banner	prince
	fake	truth	fraud	expand	fetch
	fat	broad	tiny	click	witch
	fee	wage	permit	mummy	truce
	foster	assist	aid	cash	sick
	gap	cavity	hole	whip	ward
	grain	wheat	rice	fairy	exit
	grin	teeth	glad	folder	queen
	heap	stack	gather	dwarf	quote

	hit	shield	slap	reader	prefer
	hook	sharp	trout	busy	neck
	hurdle	bounce	skip	duke	echo
	mask	hat	hood	tide	canoe
	raid	rob	troops	vase	clown
	saddle	pony	camel	angel	frown
	scan	copy	print	beak	shout
	elbow	muscle	bone	envy	loud
	shade	shadow	tree	kiss	mug
	silk	linen	shiny	cheese	rage
	slice	divide	sword	ghost	active
	smash	crush	grind	worm	virus
	tall	giant	height	code	worry
	trim	barber	beard	bag	spoon
	wool	yarn	goat	bread	foe
Non-homonym Set 2	abuse	harm	cruel	menu	chalk
	bet	luck	gamble	parent	collar
	burn	grill	heat	hint	famous
	dawn	dusk	bright	rebel	toss
	deaf	blind	noise	purse	golf
	dip	plunge	rinse	dragon	humble
	drift	wander	yacht	comedy	gun
	feast	supper	cake	smooth	horn
	fog	cloud	rain	scream	hug
	fur	fox	rabbit	chain	pill
	grasp	grab	snatch	melt	trial
	hay	farm	nest	pearl	resist
	honey	sauce	sweet	fun	rugby
	leap	runner	jump	owl	powder
	load	cargo	lorry	tour	rub
	loop	rope	shape	sniff	tribe
	peak	hill	climb	batch	bug
	pilot	sky	cabin	dirt	tape
	push	hurt	ram	rat	snack
	ritual	pray	cult	stew	honest
	rod	copper	cane	era	pillow
	smoke	vapour	oven	dairy	twin
	sour	apple	candy	bullet	weapon
	spy	agent	enemy	pale	toad

	teach	guide	learn	escape	edge
	tin	bottle	metal	sad	track
	torch	cave	lamp	speed	scalp
	void	null	valid	island	rural