Representation and processing of semantic ambiguity

Grzegorz Maciejewski

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
School of Psychology

December 2018

Intellectual Property and Publications

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 2 of the thesis has appeared in publication as follows: Maciejewski, G., & Klepousniotou, E. (2016). Relative meaning frequencies for 100 homonyms: British eDom norms. *Journal of Open Psychology Data*, *4*(1), p. e6. DOI: 10.5334/jopd.28. The candidate was responsible for conceiving and designing the study, collecting and analysing data, and writing the paper. The contribution of the other author was designing the study and writing the paper.

Acknowledgements

First and foremost, I thank Ekaterini Klepousniotou as it is very hard to imagine this thesis being completed without her guidance and hard work. I am particularly grateful for freedom in everything I did and her encouragement at difficult times. I also thank Mark Mon-Williams who offered advice whenever I needed it and helped me recognise the impact of this work.

I thank several undergraduate students for their assistance with data collection - Marketa Provodova, Rebecca Gardner, Timothy Carling (Experiment 2); Rebecca Shaw, Lauren Emina-Bouggard (Experiment 5); Florence Lister, Rebecca Gardner, Aamna Fardous, and Petya Grigorova (Experiment 6). I also thank Emily Oxley who helped me code responses from Experiment 1, Laura Arendsen and Zeynep Uludag who helped me practise EEG set-up, and Brian Scully who helped me find bugs in my EEG pre-processing scripts.

I thank the Economic and Social Research Council for their funding of my doctoral studentship (ES/J500215/1) and the University of Leeds for their funding of some of my research expenses.

Last but not least, I thank Amy Atkinson for numerous occasions when we took a day off to whinge or celebrate.

Abstract

One of the established findings in the psycholinguistic literature is that semantic ambiguity (e.g., “dog/tree bark”) slows word comprehension in neutral/ minimal context, though it is not entirely clear why this happens. Under the “semantic competition” account, this ambiguity disadvantage effect is due to competition between multiple semantic representations in the race for activation. Under the alternative “decision-making” account, it is due to decision-making difficulties in response selection. This thesis tests the two accounts by investigating in detail the ambiguity disadvantage in semantic relatedness decisions.

Chapters 2-4 concentrate on homonyms, words with multiple unrelated meanings. The findings show that the ambiguity disadvantage effect arises only when the different meanings of homonyms are of comparable frequency (e.g., “football/electric fan”), and are therefore initially activated in parallel. Critically, homonymy has this effect during semantic activation of the ambiguous word, not during response selection. This finding, in particular, refutes any idea that the ambiguity disadvantage is due to decision making in response selection.

Chapters 5 and 6 concentrate on polysemes, words with multiple related senses. The findings show that the ambiguity disadvantage effect arises for polysemes with irregular sense extension (e.g., “restaurant/website menu”), but not for polysemes with regular (e.g., “fluffy/marinated rabbit”) or figurative sense extension (e.g., “wooden/authoritative chair”). The latter two escape competition because they have only one semantic representation for the dominant sense, with rules of sense extension to derive the alternative sense on-line.

Taken together, this thesis establishes that the ambiguity disadvantage is due to semantic competition but is restricted to some forms of ambiguity only. This is because ambiguous words differ in how their meanings are represented and processed, as delineated in this work.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

HF – high-frequency

LF – low-frequency

SOA – stimulus-onset asynchrony

PDP – parallel distributed processing

ERP – event-related potential

BNC – British National Corpus

MRC – Medical Research Council

ITI – inter-trial interval

RT – response time

EEG – electroencephalogram; electroencephalographic

EOG – electro-oculogram

ICA – independent component analysis

MEG – magnetoencephalogram; magnetoencephalographic

fMRI - functional magnetic resonance imaging

LIFG – left inferior frontal gyrus

ISI – inter-stimulus interval

Chapter 1: General Introduction

1.1 Semantic Ambiguity

Word comprehension, particularly in highly skilled readers, is so fast and effortless that we do not necessarily realise the complexity of the different computations involved in making sense of what we are reading. Not only must we analyse the visual input and match it to a familiar word form, but we must also retrieve the meaning of that word and integrate it with the overall meaning of a sentence and discourse. The fact that all this information is available to us within only a few hundred milliseconds suggests that our language system relies on some form of a "mental lexicon" - a long-term memory store of the orthographic, phonological, semantic, and syntactic properties of the words we know (see Elman, 2004; MacDonald, 1997; Rastle, 2007; Woollams, 2015). While accessing this different information is a complex process in itself, it is certainly more complex for words that have multiple meanings.

The vast majority of words in all languages have multiple interpretations, such that word disambiguation is intrinsic to everyday communication (Bates, Devescovi, & Wulfeck, 2001; Rodd, Gaskell, & Marslen-Wilson, 2002). For example, the word "bank" can refer either to a financial institution or the ground alongside a river or lake. Each of these unrelated meanings of the word can also be used in different senses, or variants of the primary interpretation. The institution-related meaning can denote a supply (e.g., "a blood bank"), whereas the ground-related meaning can describe a mass of a substance (e.g., "a bank of snow"). Such semantic ambiguity (also referred to as "lexical ambiguity") is a ubiquitous property of natural language that requires readers to retrieve the different meanings of a word and then quickly commit to a single interpretation, making the ability to do so an essential component of any theory of language comprehension.

The work described in this thesis serves to develop an adequate and comprehensive account of how semantically ambiguous words are represented in the mental lexicon and processed out of context. To date, most studies of ambiguity have focussed on its impact on word recognition (e.g., Haro & Ferré, 2018; Jager & Cleland, 2016; Rodd et al., 2002), or the impact of context on meaning retrieval (e.g., Binder & Morris, 1995; Oden & Spira, 1983; Rayner, Cook, Juhasz, & Frazier, 2006). The former line of research, typically lexical decision studies[1] (for a review, see Lupker, 2007), attempts to account for ambiguity effects that arise prior to disambiguation, while the latter research, mainly sentence reading and cross-modal priming studies[2] (for a review, see Simpson, 1994), attempts to explain whether and how sentential context aids ambiguity resolution. Thus, relatively little is known regarding what processing mechanisms, unaffected by contextual bias, are involved in disambiguation itself. This thesis focuses on the processing of single words and addresses the outstanding question of how ambiguous words are comprehended in isolation or minimal context. The following section of this chapter describes different forms of ambiguity, mainly from the linguistic perspective. As demonstrated later, evidence for the differential processing of these distinct forms is critical to our

---

[1] In this task, participants decide whether a string of letters (e.g., “curgeon”) is a real word.

[2] In this task, participants usually listen to a sentence biasing the interpretation of an ambiguous word (e.g., “He found several <u>bugs</u> in his mobile phone”) and then make a lexical decision to a visual target word that instantiates one of the word’s meanings (e.g., “spy”).

understanding of the psychological underpinnings of semantic memory. Further sections bring together behavioural and computational findings that form the basis for the present work. This literature review describes ambiguity effects in word comprehension and presents implications for existing and future models of word processing.

1.2 Forms of Ambiguity

Perhaps the most salient form of ambiguity is homonymy in which a single phonological and orthographic word form[3] is associated with a number of unrelated meanings (e.g., “money/river bank”). Homonyms constitute a rather infrequent and incidental form of ambiguity (Weinreich, 1966). One crucial property of these words is that they are historically derived from distinct lexical items (Lyons, 1977). For instance, in the Online Etymology Dictionary (Harper, 2016), the word “bank” in the institution-related meaning was first documented in the late 15th century, originally denoting a money dealer’s table. The word derives from either Old Italian “banca” or Middle French “banque”, both of Germanic origin. In contrast, “bank” in the ground-related meaning dates back to the 12th century and derives from either Old English “benc” or Old Norse “banki”. Homonymy can be contrasted with polysemy in which a single word form corresponds to a number of related senses. Unlike homonyms, polysemes are characterised by shared etymology and conceptual overlap between the senses (Cruse, 1986; Goddard, 1998). Most lexicographers group the multiple senses of

---

[3] Homonyms differ from both heterophonic homographs and homophonic heterographs; the former have identical spelling but different pronunciation (e.g., “bow”), whereas the latter have identical pronunciation but different spelling (e.g., “knight” vs. “night”).

polysemes together, so that, for example, the interpretations of the word “virus” (an organism; a viral disease; a poisonous influence on one’s mind; a computer virus) are listed within a single dictionary entry. While there are important differences between living and non-living viruses, these senses are, nevertheless, closely related, both in terms of etymology and semantics.

Polysemy, or context-dependent variations in word interpretation, is the rule rather than the exception in most languages. In the case of English, support for this claim comes from analysis of Wordsmyth Dictionary entries[4] (Parks, Ray, & Bland, 1998) for 6431 words with the form frequency of five or more occurrences per million words in the MCWord Database (Medler & Binder, 2005). As shown in Figure 1.1 below, only 15.4% of the words in the dictionary are truly unambiguous and have only one sense. The majority of the words (77.9%) are polysemes and have more than two senses. Critically, while the remaining 6.7% have more than one meaning/dictionary entry, 94.2% of these homonymous words are in fact “hybrids” (Armstrong & Plaut, 2011) with distinct senses for at least some of their unrelated interpretations. These statistics, which are very similar to those reported by Rodd et al. (2002), demonstrate that ambiguous words, especially polysemes, constitute a large part of our repertoire of vocabulary.

[4] Counts of word meanings/senses in the Wordsmyth Dictionary were taken from Armstrong (2016).

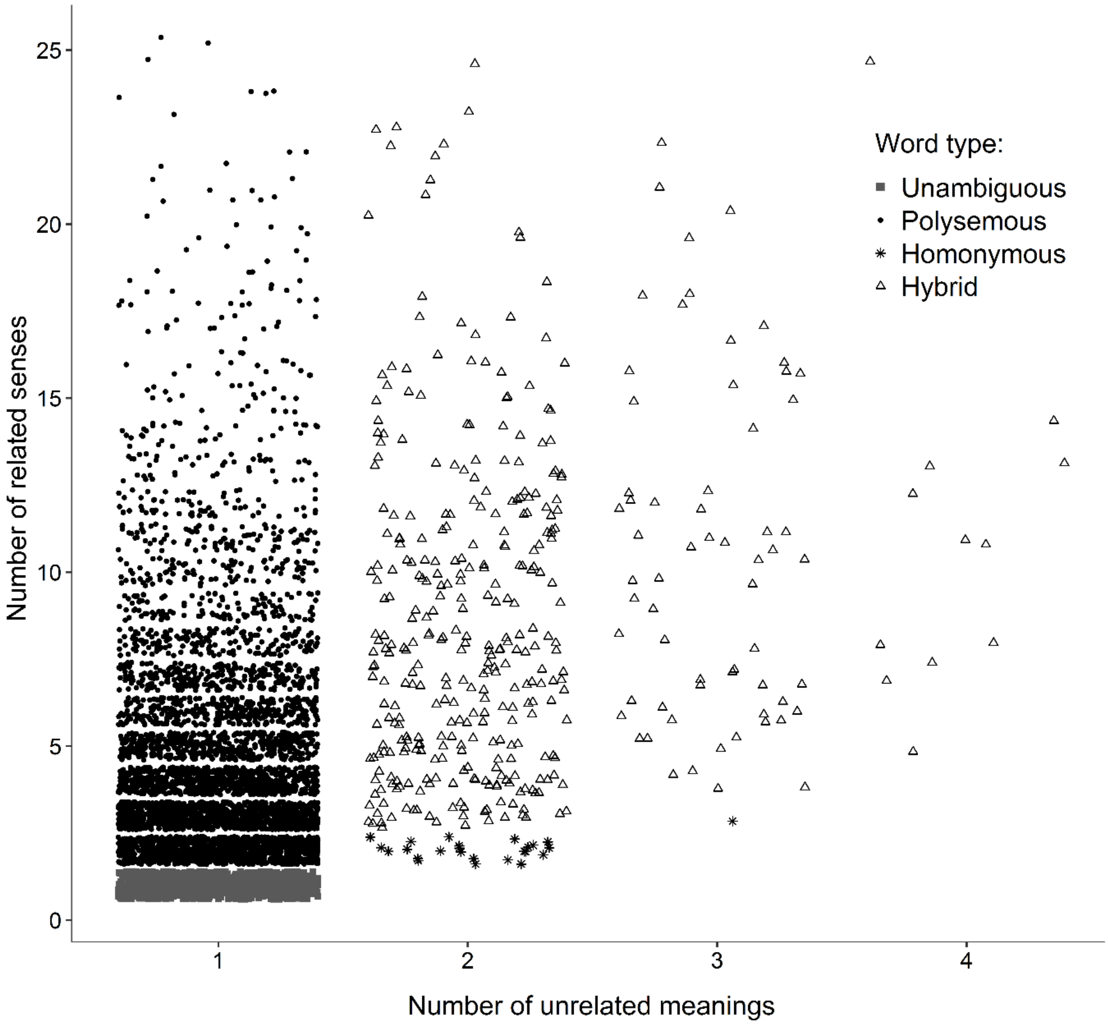

Figure 1.1: Scatterplot of the numbers of unrelated meanings and related senses for 6431 words in the Wordsmyth Dictionary (Parks et al., 1998). “Unambiguous” refers to words that have a single interpretation. “Polysemous” refers to words with one meaning but multiple senses. “Homonymous” refers to words with multiple unrelated meanings, each having only one sense. “Hybrid” refers to words with multiple unrelated meanings, with some or all having more than one sense. For the sake of clarity, two words with five meanings and 13 words with more than 25 senses are not depicted.

It is important to note that polysemes are not a homogenous group of words. Within the theoretical linguistic literature (Apresjan, 1974; Lakoff, 1987; Pustejovsky, 1991), polysemous words are traditionally divided into two forms of ambiguity that are motivated by distinct figures of speech – metaphor and metonymy. Metaphorical polysemy is based on a relation of analogy whereby one sense of a word is loosely and figuratively related to the other (e.g., “chair” as a piece of furniture or a person in charge of an organisation). Although there are some cases of more regular figurative extensions (e.g., using animal names to describe human characteristics), metaphors are largely unconstrained and variable in interpretation, such that their senses may often have obscure or even minimal conceptual overlap (Apresjan, 1974), as it is the case with the literal and the figurative interpretation of the word “twist” (“to twist an ankle” vs. “to twist the truth”).

In metonymy, on the contrary, both the primary and the secondary sense are highly related and literal (e.g., “rabbit” denoting the animal or the meat). Metonymic polysemy is by far the most common form of ambiguity that reflects language users’ tendency to use existing words to describe novel albeit conceptually related actions, concepts, and objects (Murphy, 1997; Nunberg, 1979). For instance, it is plausible that we began to use the word “rabbit” to refer to the meat soon after we had named the animal. As in many other cases of metonymy, the former sense developed from the latter through the process of sense extension (see Clark & Clark, 1979; Lehrer, 1990). Research has shown that such shifts in word interpretation are rather easy to comprehend (Clark & Gerrig, 1983; Frisson & Pickering, 2007; McElree, Frisson, & Pickering, 2006; Murphy, 2006) and come in a number of forms (Copestake & Briscoe, 1995; Pustejovsky, 1995; Rabagliati, Marcus, & Pylkkänen, 2011), including animal for meat (e.g., “rabbit”), instrument for action (e.g., “shovel”), place for people (e.g., “Edinburgh”), and producer for product alternations (e.g., “Tarantino”). In this regard, metonymic polysemy follows regular and predictable rules of sense extension that have been observed across a number of languages (Srinivasan & Rabagliati, 2015).

Taken together, this literature demonstrates that semantic ambiguity is a complex phenomenon. The multiple interpretations of words differ with respect to etymological derivation, relatedness in meaning, frequency of usage, and syntactic class to name a few. As evident in the following section, the question of whether these distinct properties of ambiguous words have any psychological implications has gradually become a crucial one.

1.3 Ambiguity Disadvantage

One may surmise that ambiguity makes word processing more difficult without any clear benefit, especially when context is absent or insufficiently constraining. This is partly true – ambiguity has either inhibitory or facilitatory influences on word processing depending on both the relatedness between the multiple interpretations of a word and what the reader must do with the word. Numerous lexical decision studies (e.g., Armstrong & Plaut, 2008; Hoffman & Woollams, 2015; Klepousniotou & Baum, 2007; Locker, Simpson, & Yates, 2003; Rodd et al., 2002) have shown that while polysemy benefits word recognition, homonymy appears to have no influence during this early stage of word processing. There is some indication of a homonymy disadvantage in lexical decisions (cf. Haro, Demestre, Boada, & Ferré, 2017; Hino, Kusunose, & Lupker, 2010), but the effect is weak and contingent on task demands, such as using pseudo-homophonic[5] non-words, that are assumed to increase the impact of semantics on orthographic processing (for a recent review on polysemy and

[5] Pseudo-homophones are non-words that share pronunciation with real words (e.g., “brane”).

homonymy effects in visual word recognition, see Eddington & Tokowicz, 2015). Unlike the lexical decision studies, the work described in this thesis focuses on the relatively under-represented investigations into ambiguity effects on the later stage of word processing - word comprehension. After all, the ultimate goal of reading is to make sense of words, rather than to simply recognise them, and it is this process that ambiguity appears to complicate the most. The following section of the chapter reviews findings from studies that employed tasks requiring word disambiguation but constrained the influence of context on the processing of homonyms and polysemes.

Research into word comprehension in the absence of contextual bias has consistently shown slower reading/response times for ambiguous than unambiguous words (Duffy, Morris, & Rayner, 1988; Frazier & Rayner, 1990; Gottlob, Goldinger, Stone, & Van Orden, 1999; Hino, Lupker, & Pexman, 2002; Jager & Cleland, 2015; Pexman, Hino, & Lupker, 2004; Piercey & Joordens, 2000; Rayner & Duffy, 1986). This so-called "ambiguity disadvantage effect" appears to have been first observed in early eye-movement studies of sentence reading. In particular, Rayner and Duffy (1986) investigated participants' reading of homonyms (e.g., "boxer") and non-homonyms (e.g., "puppy") in late-disambiguation sentences, in which context always followed the ambiguous word and supported its low-frequency (LF) meaning (e.g., "We knew the boxer/puppy was barking all night"). Their results showed that both gaze[6] and first fixation durations were significantly longer for homonyms, although this effect was restricted to homonyms with balanced meaning frequencies (e.g.,

---

[6] Gaze duration refers to the sum of all fixations on a word prior to moving to another word.

“electric/football fan”). “Unbalanced homonyms” with one high-frequency (HF) meaning (e.g., “boxer” denoting a sportsman) were, on the other hand, read as fast as their unambiguous counterparts (see also Duffy et al., 1988).

Further evidence that homonymy slows comprehension comes from studies in which participants made semantic relatedness decisions to pairs of words (Gottlob et al., 1999; Pexman et al., 2004; Piercey & Joordens, 2000). For example, Gottlob et al. (1999) reported that relatedness decisions to pairs involving homonymous primes (e.g., “hide-skin”) were slower than to those involving non-homonymous primes (e.g., “leg-limb”). Critically, this effect was observed regardless of whether participants disambiguated the homonym towards the HF or the LF meaning, and regardless of whether the ambiguous word appeared first or second in a pair. This pattern of results was later replicated by Pexman et al. (2004) who also showed that homonymy affected relatedness decision times even when the prime and the target were presented simultaneously, as opposed to the stimulus-onset asynchrony (SOA) of 1000 ms used in Gottlob et al.’s (1999) studies.

The evidence indicates that homonymy slows word comprehension, though this disadvantage effect appears to be more pronounced for words with multiple meanings of similar frequency of usage (Duffy et al., 1988; Rayner & Duffy, 1986). This finding is consistent with the broader literature on ambiguity resolution which suggests that although all meanings are activated upon reading an ambiguous word, the level and the time-course of this activation is influenced by both meaning frequency and sentential context (for a review, see Twilley & Dixon, 2000). Although a few studies showed that sufficiently strong context is able to constrain meaning activation (e.g., Glucksberg, Kreuz, & Rho, 1986;

Simpson & Krueger, 1991; Vu, Kellas, & Paul, 1998), most findings indicate that context cannot fully suppress activation of the dominant meaning (e.g., Binder & Morris, 1995; Chen & Boland, 2008; Huettig & Altmann, 2007; Peleg, Giora, & Fein, 2008; Sereno, O'Donnell, & Rayner, 2006; Tabossi & Zardon, 1993). Thus, the consensus, exemplified in the influential "reordered access" model of ambiguity resolution (Binder & Rayner, 1998; Duffy et al., 1988; Duffy, Kambe, & Rayner, 2001), is that meanings that are highly frequent or supported by surrounding context are more readily available and more likely to be selected. With respect to the study of the ambiguity disadvantage, the implication is that, in isolation or neutral context, the impact of ambiguity on word comprehension may be minimal when the LF meaning is barely activated (i.e., unbalanced homonymy), but strong when the different meanings are activated in parallel and to the same extent (i.e., balanced homonymy). This view, in line with the findings of Duffy et al. (1988) and Rayner and Duffy (1986), is discussed in more detail in Chapters 2-4.

The demonstration that homonymy incurs a processing disadvantage raises the important question of whether polysemy has a similar effect. After all, homonyms are a unique type of words that constitute a rather small part of our vocabulary (see Figure 1.1 above), and there is growing evidence to suggest that homonymy and polysemy have differential effects on word processing, at least in the early stages of word recognition (e.g., Armstrong & Plaut, 2016; Beretta, Fiorentino, & Poeppel, 2005; Klepousniotou & Baum, 2007; Rodd et al., 2002; Tamminen, Cleland, Quinlan, & Gaskell, 2006).

Early studies on polysemes set to examine the reading of ambiguous targets embedded in late-disambiguation sentences, as per Rayner and Duffy

(1986). In particular, Frazier and Rayner (1990) contrasted nouns that were homonyms with abstract and concrete meanings (e.g., “jam”), homonyms with animate and inanimate meanings (e.g., “pitcher”), metonyms with abstract and concrete senses (e.g., “dinner”), or completely unambiguous (e.g., “treasure”). As expected, the study found that participants spent more time reading the two types of homonyms than both metonyms and unambiguous words. There was, however, no indication of a processing cost for metonyms (see also Pickering & Frisson, 2001).

Unlike these eye-movement studies, further research focused on the general contrast between homonymous and polysemous words, without making a distinction between specific types of polysemy (Hoffman & Woollams, 2015; Pexman et al., 2004). For example, Pexman et al. (2004) compared relatedness decisions to homonymous and polysemous words, where the number of meanings and their semantic relatedness were established based on subjective ratings. The results showed slower responses to word pairs involving both homonyms and polysemes than to those involving unambiguous control words. Interestingly, Pexman et al. (2004) reported that this disadvantage effect was significantly greater for homonyms than polysemes.

Consistent evidence that polysemy may hinder comprehension comes from Hoffman and Woollams (2015), who determined the polysemous status of their word stimuli on the basis of semantic diversity - a corpus-based measure of the extent to which a given word appears in diverse contexts (see Hoffman, Lambon Ralph, & Rogers, 2013). The rationale was that the phenomenon of polysemy is best captured by graded contextual variation of word usage rather than participant-generated or dictionary estimates of how many related senses a word might have. In their second experiment involving a semantic relatedness decision task, Hoffman and Woollams (2015) reported that words high in semantic diversity (e.g., “paper”) were responded to more slowly than words low in semantic diversity (e.g., “cough”), in line with Pexman et al.’s (2004) finding that the ambiguity between contextually diverse, albeit related, interpretations incurs a processing cost.

In summary, tasks requiring meaning selection have shown a processing cost for ambiguous words. This ambiguity disadvantage seems to be modulated by the semantic relatedness between the multiple interpretations of a word, or the form of ambiguity. The effect is substantial for balanced homonymy (Duffy et al., 1988; Gottlob et al, 1999; Piercey & Joordens, 2000; Rayner & Duffy, 1986), moderate for loosely defined polysemy (Hoffman & Woollams, 2015; Pexman et al., 2004), and minimal for metonymy (Frazier & Rayner, 1990). Given the scarcity of research into word disambiguation out of context, it comes as no surprise that little is known as to why the ambiguity disadvantage arises, and what it reveals about the representations and processes involved in word comprehension. The following section of the chapter discusses theoretical and computational work that attempts to answer these two questions.

1.4 Semantic and Non-semantic Accounts of the Disadvantage

1.4.1 Semantic Competition

The ambiguity disadvantage in word comprehension is an inherent prediction of the “distributed” view of lexical-semantic representation (e.g., Gaskell & Marslen-Wilson, 1997; Harm & Seidenberg, 1999; Hinton & Shallice, 1991; Plaut, 1997; Seidenberg & McClelland, 1989). In short, parallel distributed processing (PDP), or connectionist, models postulate that the words we know are represented by units corresponding to their orthographic and semantic features. These units are distributed, in the sense that activation of every unit contributes to the representation of a number of words that share the same lexical or semantic features. There are connections among orthographic and semantic units which, as a result of learning processes, acquire different weights reflecting the appropriate relationship between form and meaning, such that reading first triggers activation of orthographic units which then spreads to semantic units (for more detail, see Seidenberg, 2007). Therefore, within the PDP framework, it is the weights on the connections that determine both the speed and outcome of the semantic activation process (i.e., full retrieval and selection of a word's meaning).

For unambiguous words, the orthographic pattern of activation is always associated with the same semantic pattern, which strengthens connections among the units and facilitates future form-to-meaning mapping. Ambiguity, on the contrary, precludes the development of such strong connections. The orthographic pattern for words such as "bank" is ambiguous; it is associated with a number of semantic patterns corresponding to the different word referents ("money/river bank"), which gives rise to partial activation of the multiple semantic representations, also referred to as a "blend state" (Piercey & Joordens, 2000). As semantic activation increases, the representations begin to compete for full activation (i.e., meaning selection), as only one of them can be fully activated to complete comprehension/disambiguation. According to PDP models of ambiguity (e.g., Armstrong & Plaut, 2008; Joordens & Besner, 1994; Kawamoto, Farrar, & Kello, 1994; 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 rests on the assumption that all ambiguous words have multiple semantic representations that compete during the activation process. The research discussed in the previous section has shown, however, that this is unlikely to be the case. The disadvantage effect appears to be large for words with unrelated meanings (Duffy et al., 1988; Gottlob et al., 1999; Piercey & Joordens, 2000; Rayner & Duffy, 1986) but smaller (Pexman et al., 2004) or even non-existent (Frazier & Rayner, 1990) for words with related senses. This evidence challenges the semantic competition account and presents it with a daunting task of explaining why some forms of ambiguity may not necessarily hinder word processing.

It is important to note that the effect of meaning relatedness in the ambiguity disadvantage is largely consistent with the broader literature on the processing of homonymy and polysemy (for a comprehensive review, see Chapter 5 in this thesis). In short, findings indicate that there are important differences in how the unrelated meanings and related senses of ambiguous words are represented in the mental lexicon (e.g., Beretta et al., 2005; Frazier & Rayner, 1990; Klepousniotou, Titone, & Romero, 2008; Pickering & Frisson, 2001; Rodd et al., 2002; cf. Klein & Murphy, 2001), though there is still little consensus on what exactly these differences are. Some researchers (e.g., Armstrong & Plaut, 2008; Rodd et al., 2004) posit that since the different senses of polysemes share at least some semantic features (e.g., “to dip a brush in paint” vs. “to take a dip in the pool”), they may have separate but overlapping semantic representations. Under this account, polysemes are assumed to produce less semantic competition (compared to homonyms with separate representations) that involves only those features that are unique to the different senses. In contrast, others propose that polysemes, especially those with highly related senses (i.e., metonyms), should not produce competition at all as they appear to have only one representation corresponding to the “core” features of the multiple word referents (e.g., Frazier & Rayner, 1990; Frisson & Pickering, 1999; Li & Slevc, 2017). There is also compelling evidence for representational and processing differences within polysemy (e.g., Brown, 2008; Klepousniotou, 2002; Klepousniotou & Baum, 2007; Klepousniotou et al., 2008). In particular, Klepousniotou et al.’s (2008) work suggests that the degree of competition for polysemes may depend on the overlap of their senses, such that competition does not occur for words with a single representation (i.e., metonyms with highly related senses) but has its effect for words with separate but overlapping representations (i.e., metaphors with loosely related senses). Taken together, even though there is disagreement on how exactly the related senses of words are stored in the lexicon, it is becoming increasingly clear that homonymy and polysemy differ in representation, and could therefore entail different levels of semantic competition. How the competition account can reconcile this issue is discussed in more detail in Chapter 5.

1.4.2 Decision Making

The semantic competition account proposed in the ambiguity-processing models has been challenged by Pexman et al. (2004), who argued that the ambiguity disadvantage effect reported in relatedness decision studies (Gottlob et al., 1999; Hoffman & Woollams, 2015; Pexman et al., 2004; Piercey & Joordens, 2000) reflects task-specific decision-making demands, rather than competition at the semantic level. As described above, Pexman et al. (2004; Experiments 3a & 3b) reported a substantial processing cost for word pairs involving homonyms on related trials, regardless of whether the target word referred to the HF (e.g., “hide-conceal”) or the LF meaning (e.g., “hide-skin”). Interestingly, in their follow-up experiment, Pexman et al. (2004, Experiment 3c) found that relatedness decisions to the same homonyms, paired with unrelated targets (e.g., “hide-glass”), did not significantly differ from those made to non-homonyms. Having replicated this pattern of results for polysemes (Experiments 5a, 5b, & 5c), the researchers argued that the null effect on unrelated trials challenges PDP models and their prediction of semantic competition.

Pexman et al. (2004) reasoned that if the ambiguity disadvantage were truly due to the semantic activation process, its effects would be observed both on related and unrelated trials - whenever readers needed to resolve semantic ambiguity. To accommodate their findings, the researchers posited that the disadvantage effect on related trials is a task artefact caused by response conflict during the response-selection phase of the relatedness decision task. More specifically, the claim is that homonyms (e.g., “bank-coin”) activate their multiple meanings (“money/river bank”) which then trigger conflicting responses to the target (“yes” and “no”, respectively). Therefore, the comprehension of homonyms appears to be slower because participants might take additional time to decide which word meaning should serve as response input. Critically, such response conflict is absent on unrelated trials (e.g., “bank-comet”), hence the null ambiguity effect when responding to unrelated word pairs.

Further evidence for the idea that the ambiguity disadvantage lies in decision-making processes comes from Hino, Pexman, and Lupker (2006), who contrasted homonymous, polysemous, and unambiguous words in a semantic categorisation task[7]. Since the meanings of ambiguous words (e.g., “bear”) often fall into different categories, and might therefore create response conflict similar to that in relatedness decision tasks, Hino et al. (2006) set to examine the processing of ambiguous words on “no” trials so that their meanings never fell into a category in question (e.g., “bear” in reference to the vegetable category). The results revealed a processing disadvantage for homonymous but not polysemous words. Critically, this effect of homonymy was observed only 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) argued that the observed disadvantage effect was due to decision-making processes, although different from those proposed by Pexman et al. (2004). They attributed the effect to the nature of the decision category (see also Hargreaves, Pexman, Pittman, & Goodyear, 2011; Siakaluk, Pexman, Sears, & Owen, 2007). When the category is broad (e.g., living object), participants must retrieve a large number of semantic features of the target word’s referent and decide whether either of them is true of the category. For ambiguous words, this process may take considerably longer because participants need to

---

[7] In this task, participants decide whether a word’s meaning refers to a given category (e.g., object).

retrieve and analyse the features of the different word referents, in case one of them falls into the category in question. In contrast, when the category is well-defined and narrow (e.g., vegetable), participants do not need to retrieve as many features; they can correctly categorise these words on the basis of the existence, or non-existence, of only a small number of features that are likely true of the different word referents, whilst ignoring irrelevant features that would otherwise slow comprehension. The overall argument, then, is that the impact of ambiguity arises only when decisions relevant to task performance are somewhat more difficult to make.

To recapitulate, the decision-making account has challenged a key assumption of PDP models of word processing. While these models postulate that the meanings of ambiguous words have separate semantic representations that compete for activation (e.g., Armstrong & Plaut, 2008; Borowsky & Masson, 1996; Rodd et al., 2004), both Pexman et al. (2004) and Hino et al. (2006) have suggested that this competition does not actually take place at the semantic level. Under their account, ambiguity may slow performance on semantically engaging tasks due to decision making, but not due to competitive processes involved in comprehension. However, a closer look at the evidence for this proposal reveals the need for further scrutiny. Support for the decision-making account comes only from the findings of null effects of ambiguity on “no” trials in either relatedness decision or categorisation tasks; there have been no studies that directly manipulated decision making, without changing the level of semantic activation in a task. This raises the issue of whether “no” responses are suitable for exploring ambiguity resolution.

While relatedness decisions on “yes” trials require participants to select one of the meanings of an ambiguous word in order to correctly respond to the target word (e.g., “bat-baseball”), such disambiguation is unlikely, or at least unattested, for “no” trials. Since there is no demand for commitment to a particular meaning when responding to unrelated word pairs (e.g., “bat-water”), it is possible that relatedness decisions on “no” trials do not and should not show competition effects because they are made based on shallow activation of the semantic representations of ambiguous words. Note that this novel account could also accommodate the null ambiguity effect in semantic categorisation tasks with narrow categories. As discussed above, Hino et al. (2006) contend that when the task involves a narrow category, participants do not retrieve the different features of the multiple word referents but analyse only a small set of features relevant to a category in question. Precisely the same could be said to explain the null ambiguity effect in such a task - words with multiple meanings do not slow semantic processing because the task does not require word disambiguation but rather relies on low-level activation of semantic information (for a similar view, see Armstrong & Plaut, 2016).

A closer look at the patterns of responses in Hino et al.’s (2006) studies lends support to this view. The researchers reported that the overall responses in their tasks with narrow categories were much shorter than those in the tasks with broader categories, such that the former were almost as fast as responses to the same stimuli in their lexical decision task. Although responding on trials involving narrow categories does not preclude semantic activation (see Forster & Hector, 2002; Pecher, Zeelenberg, & Wagenmakers, 2005; Rodd, 2004), it is unlikely that such a task requires complete semantic activation, and is thus sensitive to semantic properties of words (for similar views, see Forster, 2004; Landauer & Freedman, 1968). Overall, then, it remains unclear whether the evidence from “no” trials in relatedness decision and categorisation tasks can advance our understanding of whether and how semantic ambiguity affects word comprehension. After all, PDP models (e.g., Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004) predict competition to arise only when the multiple representations compete for full semantic activation (i.e., when the task requires word disambiguation).

1.5 Aims of the Thesis

In summary, a number of semantically engaging tasks, such as sentence reading and relatedness decisions, have consistently shown that semantic ambiguity slows word comprehension (e.g., Gottlob et al., 1999; Hoffman & Woollams, 2015; Piercey & Joordens, 2000; Rayner & Duffy, 1986;). The challenge of explaining this disadvantage effect has provided a strong impetus to the development of different accounts of ambiguity representation and processing. Under the competition account (e.g., Armstrong & Plaut, 2008; Borowsky & Masson, 1996; Rodd et al., 2004), the delay in comprehension arises because ambiguous words have multiple semantic representations that compete when readers need to settle on a particular meaning. Under the decision-making account (Hino et al., 2006; Pexman et al., 2004), no such competition occurs; the delay arises solely due to response-selection difficulties. However, an important caveat, discussed above, is that meaning selection and response selection are naturally related and temporally overlapping processes. That is, tasks that are low in decision-making demands, such as relatedness decisions to unrelated word pairs, may not require meaning selection being a prerequisite for word comprehension. Conversely, tasks that do require meaning selection, such as deciding whether the word “bear” refers to an animal, may entail additional decision making that slows processing. Therefore, it is clear that further research must distinguish between the two mechanisms by explicitly manipulating factors specific to one but not the other.

The work presented in this thesis examines the ambiguity disadvantage in relatedness decisions (i.e., decisions on whether the ambiguous prime and the unambiguous target are related in meaning). As noted earlier, the task requires disambiguation but eliminates the impact of contextual bias, and has thus the best potential to uncover the key principles underlying the comprehension of ambiguous words. The primary aim is to establish the locus of the disadvantage by manipulating properties of ambiguous words that influence semantic activation but not decision-making processes – namely, meaning frequency (Chapters 2-4) and meaning relatedness (Chapters 5 & 6).

The question of why the ambiguity disadvantage arises has important implications for our understanding of the mental lexicon. Within the PDP framework, finding a single meaningful interpretation of an ambiguous word entails an additional process of deactivating competitors, and this prediction of semantic competition is inherent to each and every model with distributed lexical-semantic representation. Evidence that such competition does not take place would call into question not only the aforementioned PDP models of ambiguity processing (e.g., Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004), but also influential models of visual and spoken word recognition that are based on the same postulates of representation (e.g., Gaskell & Marslen-Wilson, 1997; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Van Orden, Pennington, & Stone, 1990).

Indeed, Hino et al. (2006) and Pexman et al.'s (2004) work suggests that the prediction of semantic competition in PDP models is incorrect – activation of each of the semantic representations of ambiguous words arises independently of the other, as in early localist models of ambiguity (Jastrzembski, 1981; Rubenstein, Garfield, & Millikan, 1970). Pexman et al. (2004), in particular, argue that such an explanation would hold true if a model 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 "bank" is represented within one semantic space, whereas the river-related meaning in another. Overall, then, it is important that we establish the precise cause of the ambiguity disadvantage before we make any further inferences about the structure of the mental lexicon. Whatever such research uncovers, any general account of how our knowledge of words and their meanings is represented and accessed must be able to accommodate this finding.

The second aim of the current work is to investigate whether all forms of ambiguity incur a processing disadvantage. As reviewed above, preliminary research on the role of meaning relatedness has provided inconsistent results. Pexman et al. (2004) and Hoffman and Woollams (2015) found that, like homonyms, polysemes hinder semantic processing, while Frazier and Rayner (1990) and Hino et al. (2006) reported no such evidence. Extending this line of research, the experiments reported in Chapters 5 and 6 not only contrast the impact of homonymy and polysemy in word comprehension, but they also investigate the different forms polysemy can take, given the growing evidence that the linguistic distinction within polysemy may be psychologically valid (e.g., Brocher, Koenig, Mauner, & Foraker, 2018; Klepousniotou & Baum, 2007; Lopukhina, Laurinavichyute, Lopukhin, & Dragoy, 2018). Thus, the question of whether meaning relatedness matters for word comprehension, as it seems to do for word recognition (for a review, see Eddington & Tokowicz, 2015), is essential in order to develop a comprehensive and adequate account of how ambiguous words are represented and processed. While existing models (e.g., Kawamoto, 1994; Rodd et al., 2004) assume that all ambiguous words have multiple semantic representations that compete with one another in the race for semantic activation, it is plausible that this is not the case.

Chapter 2: UK-based Norms of Meaning Frequency

2.1 Introduction

Following the evidence, reviewed in Chapter 1, that semantic ambiguity slows word comprehension, the first set of the present experiments investigates why this disadvantage occurs, and what it reveals about the representation and processing of semantic information. These experiments focus on homonyms, expecting that if any form of ambiguity entailed semantic competition, it would be foremost observed for homonyms whose completely unrelated meanings are unanimously assumed to have separate representations (for a review, see Eddington & Tokowicz, 2015). To recapitulate, under the semantic competition account (e.g., Armstrong & Plaut, 2008; Borowsky & Masson, 1996; Kawamoto, 1993; Rodd et al., 2004), homonyms have distributed semantic representations whose initial parallel activation produces competition for further activation. In contrast, under the decision-making account (Hino et al., 2006; Pexman et al., 2004), this activation does not produce competition but gives rise to effortful decision-making processes, such as response-conflict resolution. What the two proposals have in common, then, is the implicit assumption of parallel meaning activation. The evidence reviewed below shows, however, that this assumption is at odds with existing evidence (for a review, see Twilley & Dixon, 2000), and suggests that understanding how activation of word meanings arises is critical to explaining the disadvantage effect.

2.1.1 Meaning Frequency

Investigations into meaning activation have shown that readers do not retrieve the different meanings of homonyms simultaneously, contrary to the postulate of the semantic competition and the decision-making account (for a detailed review, see Twilley & Dixon, 2000). As noted in Chapter 1, whenever ambiguous words are encountered in isolation or minimal context, relative meaning frequency exerts a substantial impact on both the time-course and the level of meaning activation. For example, in their event-related potential (ERP) study, Klepousniotou, Pike, Steinhauer, and Gracco (2012) used homonyms to prime lexical decisions to targets related to either the HF or the LF meaning of the ambiguous words. Focusing on the N400 component, considered to be an index of semantic processing (for a review, see Kutas & Federmeier, 2011), the results revealed significant priming effects (i.e., reduced N400 amplitudes relative to unrelated trials) for both HF- and LF-meaning targets at the 250-ms SOA. The effect was, however, stronger for the former; activation of the HF meaning involved a larger network in the right hemisphere, which seems to indicate, as Klepousniotou et al. (2012) suggested, that participants accessed more semantic features of the HF than the LF word referent (see also Meade & Coch, 2017).

Simpson and Burgess (1985) also used homonymous words to prime lexical decisions to HF- and LF-meaning targets but varied the SOA. Their study revealed an interesting pattern of results. Simpson and Burgess (1985) reported a significant priming effect for HF-meaning targets at the 16-ms SOA. At the 100-ms SOA, there was also a priming effect for LF-meaning targets, albeit a considerably smaller one. At the 300-ms SOA, priming of the LF meaning began to gradually decrease, while that of the HF meaning increased further (see also Burgess & Simpson, 1988; Frost & Bentin, 1992; Simpson & Krueger, 1991). The overall implication from the studies is that the speed and the level of activation of a particular word meaning is strongly determined by its frequency. In fact, activation of the dominant meaning is so strong that even inconsistent context has been shown to be unable to fully suppress it (e.g., Chen & Boland, 2008; Peleg et al., 2008; Sereno et al., 2006; Swaab, Brown, & Hagoort, 2003; Tabossi, 1988). This indicates that any adequate account of how activation of multiple meanings affects comprehension must recognise the influence of both meaning frequency and time on the activation process.

Drawing on this literature, it seems plausible that the degree of semantic competition involved in word comprehension may be strongly determined by relative meaning frequency. For homonyms with balanced meaning frequencies (e.g., “football/electric fan”), semantic competition is, in principle, maximal. In the absence of context, the multiple semantic representations are initially activated in parallel and to the same extent, as neither meaning frequency nor context can bias the activation process. For the unbalanced counterparts (e.g., “money/river bank”), the impact of meaning frequency on semantic activation eliminates, or at least reduces, competition. Readers may fully retrieve the HF meaning very fast, such that the alternative meaning does not reach a sufficient level of activation to compete in the race for further activation. This proposal on the role of meaning frequency in semantic competition is further discussed and explored in Chapters 3 and 4.

2.1.3 Norms of Meaning Frequency

The main aim of Experiment 1 was to collect meaning-frequency ratings for a large set of homonymous words to be used in further experiments contrasting the processing of balanced and unbalanced homonymy (reported in Chapters 3 & 4). There are a few such norms in both Canadian and American English. These early norming studies (e.g., Geis & Winograd, 1974; Nelson, McEvoy, Walling, & Wheeler, 1980; Twilley, Dixon, Taylor, & Clark, 1994) typically used a word association task whereby the relative frequency of a word meaning was determined based on the proportion of participant-generated associates relating to that meaning (e.g., “coin” in response to “bank”). Although these norms have been demonstrated to provide fairly reliable estimates of meaning dominance (for a review, see Twilley et al., 1994), there is no evidence to suggest that these estimates are equally reliable across the distinct dialects of the English language, and as such they seem unsuitable for examining the effect of meaning frequency in British-English speakers recruited in the current experiments. Experiment 1, thus, provides the first norms of meaning frequency in British English and investigates dialectal variation in native speakers’ long-term experience with the different meanings of ambiguous words.

At first glance, the idea that dialect shapes our experience with different word meanings, just as it appears to shape our experience with different word forms (for tentative evidence, see Armstrong, Zugarramurdi, Cabana, Lisboa, & Plaut, 2015), has much merit. If such between-dialects variation exists, it is likely due to various socio-cultural factors that influence particular instances of words, as opposed to the entire repertoire of vocabulary. For instance, the campsite-related meaning of “camp” may be more common in American than British English simply because this type of leisure is far less common in the UK. While this factor might account for dialectal differences in meaning frequencies for several other ambiguous words, it is unlikely that all differences follow such a pattern. Preliminary evidence for dialectal variation in meaning dominance comes from a recent norming study that compared meaning-frequency ratings from European- and Rioplatense-Spanish speakers (Armstrong et al., 2015). The results revealed fairly substantial variation (R^2 = .72) between the dialects, but, critically, no hints of a systematic pattern underlying these differences. In order to extend Armstrong et al.’s (2015) work on dialectal variation, Experiment 1 provides the first norms in British English and compares them to analogous norms in American English (Armstrong, Tokowicz, & Plaut, 2012).

Meaning-frequency ratings in the current experiment were collected online using the eDom norming method that was introduced by Armstrong et al. (2012). Unlike the word association task used in early norming studies (e.g., Geis & Winograd, 1974; Nelson et al., 1980), this procedure relies on explicit participant ratings of frequency. Raters are presented with dictionary definitions of the meanings of homonyms and estimate their relative frequencies (as a percentage score) on the basis of their personal experience. Experiment 1 used the eDom procedure mainly because it allowed for the direct comparison of the current ratings in British English to those in American English (Armstrong et al., 2012). Furthermore, Armstrong et al. (2012) presented compelling evidence that, compared to word association norms (Twilley et al., 1994), the eDom method provides more accurate estimates of meaning dominance that can account for greater variance in homonym processing, at least with regards to lexical decision performance.

In summary, Experiment 1 provides the first UK-based norms of meaning frequency for a large number of homonyms that were carefully selected and validated for further studies on the processing of balanced and unbalanced homonyms. A subsidiary aim is to establish whether and how native speakers' experience with the different meanings of the homonyms varies across dialects of the English language.

2.2 Method

2.2.1 Participants

One hundred monolingual native speakers of British English [55 females; aged 19-39 (*M* = 28.1, *SD* = 5.3)] participated in the experiment. All participants were students (*n* = 30) or professionals (*n* = 70) born and resident in the UK. Individuals with any language-related difficulties or disorders, or those with an education qualification below A Level[8] did not take part[9]. All participants were recruited via Prolific (http://prolific.ac/), completed the task online via Qualtrics (http://qualtrics.com)[10], and received £3 for their time. The experiment received

[8] A UK qualification taken by school students aged 16-18.

[9] These two criteria applied to every stimulus pre-testing study reported in this thesis.

[10] All stimulus pre-testing studies described throughout the thesis were conducted online using Qualtrics surveys (http://qualtrics.com). All participants in these studies were recruited via Prolific Academic (http://prolific.ac) and were paid based on an hourly rate of £6.

ethical approval from the School of Psychology, University of Leeds Ethics Committee.

2.2.2 Stimuli

One hundred words with the same spelling and pronunciation were selected from a large list of homonymous words in Armstrong et al. (2012). Ninety-three of them had two separate entries in the Wordsmyth Dictionary (Parks et al., 1998) being suggestive of homonymy. In a few cases (*n* = 7), the dictionary listed a third but highly uncommon meaning of the word (e.g., “sack” denoting a light-coloured dry sherry). Most of the words had either noun-noun (*n* = 47) or noun-verb (*n* = 36) interpretations. The remaining words had noun-adjective (*n* = 9), verb-adjective (*n* = 5) and verb-verb (*n* = 3) interpretations. Stimulus selection excluded homonyms for which one of the two meanings was highly uncommon (e.g., “frail” denoting a basket made of dried rushes), known only to a specific population (e.g., “bleak” denoting a type of a fish), archaic (e.g., “burden” denoting the refrain of a song), or exclusively used by British-English (e.g., “chap” denoting a man) or American-English speakers (e.g., “bus” denoting to clear restaurant tables). The latter criterion, applied based on the information on dialectal differences in word usage in the Oxford Dictionary (Simpson & Weiner, 1989), aimed to allow for the current set of carefully selected homonymous words to be used by UK- and USA-based researchers. The set also excluded items that were ambiguous because the word form was an abbreviation (e.g., “log”) or past simple/participle (e.g., “dove”). Stimulus selection was further constrained by word length (3-6 letters) and word-form frequency (4-60 occurrences per million) in the British National Corpus (BNC, 2007). The words were also fairly homogenous with respect to 14 lexical and semantic variables, such as word-form frequency, length, and imageability (see stimulus properties in Table 2.1 below). See Appendix 1 for the list of the words and their definitions.

Table 2.1: Experiment 1: Properties of the normed homonyms.

Lexical variable	Mean (SD)	Semantic variable	Mean (SD)
Letters	4.3 (0.9)	Senses[1]	8.4 (4.2)
Phonemes	3.6 (0.8)	Senses[2]	8.8 (5.2)
Syllables	1.1 (0.4)	Imageability	4.8 (1.0)
Raw frequency	20.7 (14.0)	Concreteness	5.2 (0.8)
Log frequency	1.2 (0.3)	Age of acquisition	7.1 (2.2)
Orthographic neighbours	8.0 (5.9)	Semantic diversity	1.6 (0.2)
Log bigram frequency	2.8 (0.4)	Subjective familiarity	5.0 (0.6)

Note: Standard deviations are given in the parentheses. Form frequency, bigram frequency, and the number of orthographic neighbours come from the BNC (2007). "Senses[1]" and "Senses[2]" refer to the number of word senses in the Wordsmyth (Parks et al., 1998) and the WordNet Dictionary (Fellbaum, 1998), respectively. Semantic diversity values come from Hoffman et al. (2013). Age-of-acquisition ratings come from Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012). Imageability, concreteness, and familiarity ratings come from the Medical Research Centre (MRC) Psycholinguistic Database (Coltheart, 1981). Throughout the thesis, all variables, except for the number of senses, concreteness, age of acquisition, and semantic diversity, were obtained using N-watch (Davis, 2005).

The homonymous status of the word stimuli was confirmed based on ratings of meaning relatedness from 30 monolingual native British-English speakers [16 females; aged 19-38 (M = 29.9, SD = 4.8)] who did not take part in the norming experiment. In this pre-test, participants read the definitions of the two meanings of an ambiguous word (all taken from the Wordsmyth Dictionary, Parks et al., 1998) and rated their semantic relatedness on a 7-point Likert scale (where 1 denoted "highly unrelated" and 7 denoted "highly related"). The homonyms were rated along with 100 polysemes taken from Klepousniotou et al. (2012) and Rodd et al. (2002). For these fillers, participants rated the relatedness between two different senses of the word (e.g., “maple” denoting either a tree or the wood). As expected, the average rating was significantly lower for homonymous (*M* = 2.3, *SD* = 0.9) than polysemous words [*M* = 5.1, *SD* = 1.2; $t(99) = 18.3$, $p < .001$]. However, the data (see Appendix 1) also revealed questionable unrelatedness in meaning for eight out of the 100 words (e.g., the verb vs. the noun interpretation of “mount”) that received an average rating at or above 4 (“neither related nor unrelated”). These items were included in the norms but excluded from further experiments.

2.2.3 Procedure

The words were normed online using Qualtrics (http://qualtrics.com). As in Armstrong et al. (2012), participants estimated, as a percentage, how often each of the meanings of a homonym was implied when they encountered the word. Participants rated one word at a time in pseudo-randomised order. They first saw a homonym and indicated whether they knew the word. Wordsmyth Dictionary (Parks et al., 1998) definitions of the two meanings were shown in a panel below the word in pseudo-random order. Participants were instructed that the order and length of the definitions did not reflect the frequencies of the word meanings. They had an opportunity to list up to two additional meanings if the presented definitions were not exhaustive according to their knowledge of the word. Two text-entry boxes were provided for this purpose. If participants knew more than two additional word meanings, they were asked to list only those encountered the most. Finally, participants rated the relative frequencies of all the meanings (those in the dictionary definitions and those that they may have added themselves) using percentage scores (0-100), which had to sum up to 100 across all the meanings. On average, the experiment lasted for 40 minutes.

2.3 Results

2.3.1 Number of Word Meanings

Analysis excluded 0.3% of the responses where participants indicated that they did not know a given word. These "null responses" appeared across 12 participants and 20 words. The highest number of null responses was seven (both per participant and item), which did not warrant any data deletion and demonstrated that the group was of a suitable linguistic background. On average, participants provided 12.6 (*SD* = 11.4) additional meanings, which is suggestive of their thorough approach to the task. These "additional meaning responses" appeared across 91 words and constituted, on average, 32.0% (*SD* = 15.6) of participants' encounters with a given word. Most of the generated definitions seemed to pertain to common variants of the interpretations (e.g., "squash" denoting a sport) that were not explicitly conveyed in the presented definitions ("to press, beat, or crush into a pulp").

Further analysis established whether the additional meanings referred to a third unrelated meaning of a word or a related sense of either of the presented meanings. In the latter case, the frequency rating of an additional meaning (e.g., "mate" as a friend) was added to the rating of the related presented meaning ("mate" as a marriage partner). The researcher (GM) and another PhD student (EO), who was not involved in any other stages of the research, independently coded all of the 1,263 participant-generated meanings. One additional meaning response was deleted from the norms due to an insufficient level of detail. Each response was coded as either unrelated to either of the presented meanings or highly related to the first/second meaning of the word. There was substantial inter-rater agreement on the relatedness between the additional and presented meanings ($\kappa = .73$, $p < .001$); the raters disagreed only on 14.3% of the responses that mostly pertained to idiosyncratic definitions of the words, and the disagreement was resolved by the researcher's supervisor (EK). Nearly half (49.3%) of the additional meanings, with an average meaning-frequency rating of 25.8% (SD = 16.8), were considered unrelated to either of the presented meanings. The researcher (GM) reviewed the definitions and found consistent meanings (listed by at least five participants) for 21 of the homonyms. For these words, additional meanings were listed by 5-60 participants (M = 21.4, SD = 16.8) and had a mean meaning frequency of 37.8% (SD = 16.3).

2.3.2 Meaning Dominance

Meaning dominance for each homonym (see Figure 2.1 below) was calculated using a formal measure of meaning dominance (β) introduced by Armstrong et al. (2012). β-values range from 0 to 1, where the latter represents words with highly unbalanced meaning frequencies. The values were calculated by subtracting the rating of the less frequent meaning from the rating of the more frequent meaning and then dividing the result of the subtraction by the rating of the more frequent meaning (for these ratings and β-values, see Appendix 1).

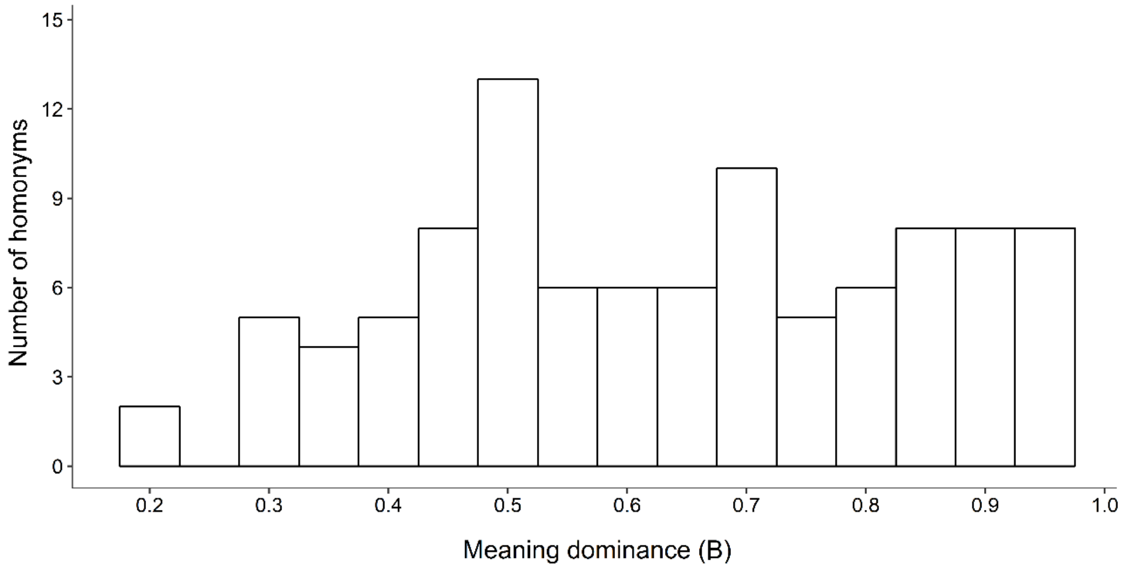

Figure 2.1: Experiment 1: Distribution of meaning dominance (β) scores. Higher β-values indicate homonyms with unbalanced meaning frequencies.

To examine between-raters variation in the estimates of dominance, participants' β-values were correlated with the group means of β across all the words, separately for each participant. Spearman's rank correlation coefficients (r_s) ranged from .02 to .85 (M = .67, SE = .02, N = 100, mean R^2 = .45), which indicated moderate inter-rater consistency in the ratings[11] being similar to those reported in the American-English (mean R^2 = .49; Armstrong et al., 2012) and European-Spanish eDom norms (mean R^2 = .48; Armstrong et al., 2015).

2.3.3 Dialectal Differences

Turning to dialectal variation in meaning dominance, analysis compared the current and analogous American-English (Armstrong et al., 2012) ratings for the first meaning of each homonym in the Wordsmyth Dictionary (Parks et al.,

---

[11] Supplementary analyses showed that participants' ratings did not significantly differ from those of the group depending on age, employment status, education level, or geographical location.

1998). This comparison was not initially possible due to different approaches to the categorisation of additional meaning responses in the two experiments. Armstrong et al. (2012) added the frequency ratings of additional meanings to the ratings of the main meanings presented in their experiment if the two were listed as senses in the Wordsmyth Dictionary (Parks et al., 1998). In the current experiment, on the other hand, the ratings of these meanings were summed only if there was substantial conceptual overlap between them, as assessed by two raters. This resulted in large numerical differences in the ratings for 15 words with obscure semantic relatedness between the participant-generated and presented meanings. For instance, in Armstrong et al.'s (2012) norms, the dominant meaning of "plane" (a flat surface) has a frequency rating of 74% because it includes the rating of a common additional meaning (an airplane) that is listed as a related sense in the Wordsmyth Dictionary (Parks et al., 1998). In the current experiment, these meanings were considered unrelated, hence the frequency rating of the surface-related meaning of "plane" (34%) was provided separately from that of the vehicle-related meaning (66%). However, for the sake of the between-dialects comparison, the ratings for such words (n = 15) were adjusted using the approach in Armstrong et al. (2012). The analysis revealed relatively large variation (R^2 = .69) in meaning dominance between the two norms (see Figure 2.2 below).

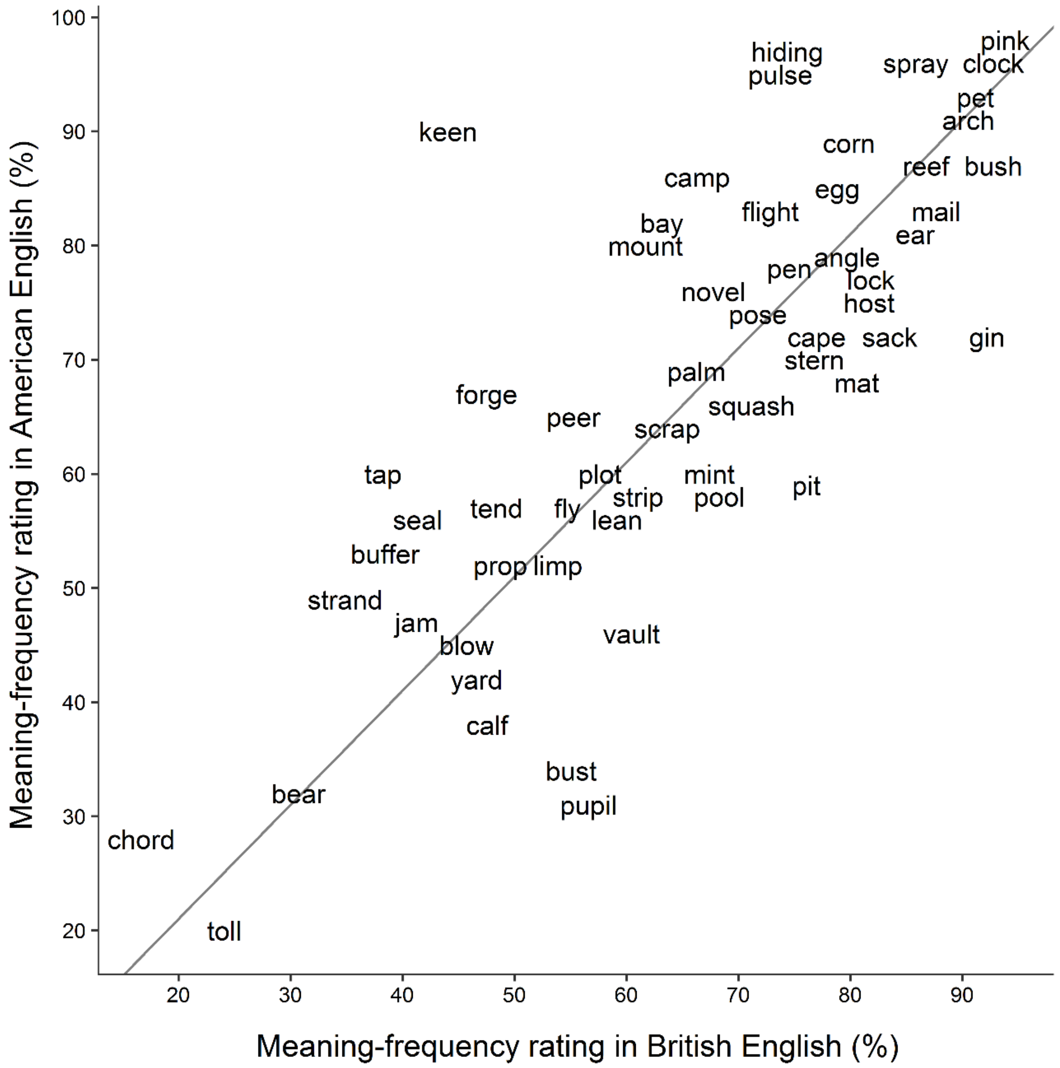

Figure 2.2: Experiment 1: Comparison of adjusted British- and American-English (Armstrong et al., 2012) frequency ratings for the first meaning in the Wordsmyth Dictionary (Parks et al., 1998). Words further from the theoretical line of best fit (e.g., "pupil") indicate larger dialectal differences between the norms. For the sake of clarity, the scatterplot does not show 40 of the 100 homonyms that would otherwise overlap with the displayed words.

2.4 Discussion

Experiment 1 provides the first UK-based norms of meaning frequencies for homonyms (available at http://osf.io/7k3eh/). These norms come from a large and diverse group of participants, and as such are representative of native British-English speakers and their different linguistic experience. Indeed, while the eDom procedure can reliably determine whether a given homonym is balanced or unbalanced, analyses of between-raters consistency demonstrated that the precise estimate of meaning dominance varies across individuals. Based on similar results, Armstrong et al. (2012, 2015) conjectured that eDom ratings contain some degree of “random noise” that lowers their inter-individual reliability, which is reasonable given the difficulty and meta-linguistic demands involved in estimating lifelong exposure to the different meanings of a word and assigning percentage scores that precisely capture that experience.

2.4.1 Variation Between Speakers

It seems reasonable to consider the idea that random noise is, at least partly, due to inherent and unsystematic differences between native speakers in their actual experience with the meanings of ambiguous words. This idea is supported by the magnitude of the between-speakers variation in eDom studies. Recent studies of word-meaning priming (Betts, Gilbert, Cai, Okedara, & Rodd, 2018; Cai et al., 2017; Davies, Porretta, Koleva, & Klepousniotou, 2017; Gaskell, Cairney, & Rodd, 2019; Gilbert, Davis, Gaskell, & Rodd, 2018; Rodd, Lopez Cutrin, Kirsch, Millar, & Davis, 2013; Rodd et al., 2016) have shown that meaning dominance is a fluid property of lexical-semantic representations that adapts to individuals’ linguistic experience and environment. For example, while the word “pupil” is generally considered a balanced homonym, exposure to its different meanings, and hence frequency ratings, would certainly differ between ophthalmologists and school teachers. Such an explanation for the inter-individual variation in meaning dominance is particularly plausible for the current norms given the diversity of the recruited participants with respect to age, education, employment, and regional British-English variation.

2.4.2 Variation Between Dialects

The main aim of Experiment 1 was to collect meaning-frequency ratings for a large set of homonyms in British English. As discussed in the introduction to the experiment, there are currently no such norms - UK-based researchers examining ambiguity processing must collect their own small-scale ratings prior to experimental testing or use those obtained in other English dialects (e.g., MacGregor, Bouwsema, & Klepousniotou, 2015; Sereno et al., 2006; Vitello, Warren, Devlin, & Rodd, 2014). The results of Experiment 1 demonstrate that establishing meaning frequency in British English was indeed warranted, and further support the idea that examining the impact of meaning dominance on the processing of ambiguity requires one to determine this variable on the basis of normative data from the corresponding dialect. Experiment 1 revealed fairly considerable variation (R^2 = .69) between the British-English and American-English ratings of meaning frequency, extending Armstrong et al.'s (2015) work on the role of dialect in English.

This variation in the ratings of meaning frequency appears to reflect genuine differences in how native British and American speakers use and encounter these words. Inspection of the words with large between-dialects differences revealed no specific pattern; there was a small number of words (n = 16) whose meanings were substantially more common in one dialect than in the other. For instance, the student-related meaning of "pupil" appears to be more frequent for British-English speakers (57% vs. 31%), whereas the campsite-related meaning of “camp” is highly dominant in American but not in British English (86% vs. 66%). Such differences in meaning dominance for only a few of the homonyms may reflect a number of cultural and linguistic factors (e.g., using synonyms to avoid ambiguity, prevalence of the word referent) that are specific to these particular words, rather than to the entire dialect. Research into these factors and their contribution to dialectal differences in how words and their meanings are used is clearly worth further scrutiny.

2.4.3 Conclusions

In summary, Experiment 1 provides researchers with a carefully selected and validated set of homonyms that can be used to explore various aspects of ambiguity processing. The present experiment appears to be the first to have determined the ambiguous status of the word stimuli based on linguistic criteria, dictionary entries, and subjective ratings. Previous studies (e.g., Armstrong et al., 2012; Azuma & Van Orden, 1997; Haro et al., 2017; Klepousniotou, 2002; Rodd et al., 2002) have used either of the methods, even though there is still no consensus as to which of them captures the nature of homonymy best (see Haro & Ferré, 2017; Lin & Ahrens, 2005). Having used all three measures, the experiment established how common and semantically unrelated the different meanings of homonyms are, and how many of the meanings native speakers actually know, providing future studies, including those reported in the following chapters, with a high-quality set of balanced and unbalanced homonyms.

Chapter 3: Semantic Competition in Word Comprehension: Evidence from Homonymy

3.1 Introduction

The work reported in this chapter examines the processing of balanced and unbalanced homonyms in order to establish whether the disadvantage effect arises due to semantic competition or decision making. The literature discussed previously suggests that the relative frequency of a word meaning determines both the time-course and level of its activation (for a review, see Twilley & Dixon, 2000), hence the proposal that the impact of ambiguity on word processing may depend on meaning frequency is not entirely new. There have been several studies that either controlled for (e.g., Armstrong & Plaut, 2016; Mirman, Strauss, Dixon, & Magnuson, 2010) or manipulated this property (e.g., Brocher et al., 2018; Klepousniotou et al., 2012; MacGregor et al., 2015). In particular, Armstrong et al. (2012) suggested that the impact of homonymy in word recognition varies depending 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, Garnett, Malyutina, & den Ouden, 2014; Klepousniotou & Baum, 2007).

It appears that meaning frequency and its impact on semantic activation may also be relevant to investigations into the ambiguity disadvantage. For unbalanced homonyms (e.g., “money/river bank”), the activation process is strongly biased toward the dominant, HF meaning, such that the alternative meaning may not even get a chance to compete for activation. For balanced homonyms (e.g., “football/electric fan”), parallel activation of their semantic representations should, on the other hand, result in considerable competition in the race for further activation. Indeed, there is already some evidence to suggest that meaning frequency may modulate the degree of competition. As mentioned in Chapter 1, studies of sentence reading (Duffy et al., 1988; Rayner & Duffy, 1986) found that the disadvantage effect is restricted to homonyms with equally frequent interpretations. Likewise, Kawamoto’s (1993) PDP model simulations also predict the slowing in comprehension to be more pronounced for balanced than unbalanced homonyms, further delineating the important role of meaning frequency in the processing of ambiguous words.

Taken together, the main premise of the present work was that if the disadvantage effect is due to competition during semantic activation, it should be sensitive to meaning frequency and its influence on semantic activation. The aim was to replicate and further investigate the finding of a disadvantage effect for balanced but not unbalanced homonyms. Experiments 2 and 3 contrasted the two types of homonyms using a semantic relatedness decision task. Unlike the eye-movement studies (Duffy et al., 1988; Rayner & Duffy, 1986), in which contextual information following an ambiguous target always supported its LF meaning, the task required participants to disambiguate each homonym toward both the HF and the LF meaning, providing a more detailed insight into word-meaning processing.

As in Pexman et al. (2004), Experiments 2 and 3 explored the processing of homonyms on both related and unrelated trials. On related trials, the different meanings of ambiguous primes might indeed trigger conflicting responses to targets (e.g., “fan-breeze”), hence any slowing in relatedness decisions could be due to decision-making, rather than semantic, processes. In contrast, on unrelated trials, all the meanings of ambiguous prime trigger the same response (e.g., “fan-snake”), eliminating response conflict and any decision making it might entail, hence any slowing in relatedness decisions on unrelated trials could only be attributed to semantic processes. In other words, if the processing disadvantage for balanced homonyms is due to competition at the semantic level, its effect should be observed on both related and unrelated trials – whenever readers encounter these words. This finding would suggest that the null ambiguity effect on unrelated trials in Pexman et al. (2004), which their decision-making account rests on, appeared because they had not considered the role of meaning frequency. Specifically, their work did not recognise that balanced and unbalanced homonyms differ in how their multiple meanings are activated, and may therefore also differ in their impact on word comprehension.

3.2 Experiment 2: Short Prime Duration

3.2.1 Method

3.2.1.1 Participants

Thirty-five students and members of staff [30 females, aged 19-35 (*M* = 25.8, *SD* = 4.9)] from the University of Leeds participated in the experiment in exchange for £3. Participants were monolingual native speakers of British English with no known history of language- or vision-related difficulties or disorders. They were all were right-handed, as confirmed using the Briggs-Nebes (1975) modified version of Annett's (1967) Handedness Inventory[12]. The experiment received ethical approval from the School of Psychology, University of Leeds Ethics Committee.

3.2.1.2 Stimuli

Fifty-six words were selected from the norms of Experiment 1. Twenty-eight of these items were unbalanced homonyms (e.g., “fan”) with the highest scores of meaning dominance, while the other half were balanced homonyms (e.g., “pen”) with the lowest scores of meaning dominance. For the former, the set excluded highly unbalanced homonyms whose alternative, LF meaning was unknown (i.e., given a frequency rating of 0%) to more than a quarter of the participants (i.e., 25 participants) in Experiment 1. This was necessary to ensure that the stimuli were truly ambiguous and processed as such. All items had low meaning-relatedness ratings (see Table 3.1 below), but three of the unbalanced and four of the balanced homonyms had a third LF meaning (e.g., the music-related meaning of “jam”). The prime set also included 56 non-homonymous control words that had only one entry/meaning in the Wordsmyth Dictionary (Parks et al., 1998). Most of these word stimuli served as filler items in the meaning-relatedness pre-test carried out as part of Experiment 1. The ratings confirmed the non-homonymous status of the primes, and there was no indication in the Wordsmyth Dictionary (Parks et al., 1998) that any of the remaining items had multiple unrelated meanings. Non-homonyms were split into two sets of 28.

---

[12] The inventory was used in all lab-based experiments.

Table 3.1: Experiment 2: Properties of the homonymous and non-homonymous prime words.

Variable	Homonymous prime		Non-homonymous prime	
	Balanced	Unbalanced	Set 1	Set 2
Example	"fan"	"pen"	"crew"	"dawn"
Meaning relatedness	2.1 (0.5)	2.0 (0.5)	5.8 (0.6)	5.8 (0.5)
Meaning dominance (β)	0.4 (0.1)	0.7 (0.1)	-	-
Letters	4.3 (0.8)	4.0 (0.8)	4.3 (0.9)	4.1 (0.9)
Phonemes	3.6 (0.7)	3.4 (0.8)	3.6 (0.7)	3.5 (0.8)
Syllables	1.1 (0.4)	1.1 (0.3)	1.1 (0.4)	1.2 (0.4)
Raw frequency	18.9 (13.4)	18.2 (10.8)	18.8 (12.4)	18.7 (10.2)
Log frequency	1.2 (0.3)	1.2 (0.3)	1.2 (0.3)	1.2 (0.2)
Orthographic neighbours	7.3 (6.0)	9.2 (6.2)	7.6 (5.0)	7.3 (5.7)
Log bigram frequency	2.8 (0.5)	2.8 (0.4)	2.8 (0.4)	2.7 (0.5)
Senses[1]	8.1 (4.3)	8.0 (3.4)	8.3 (4.0)	7.9 (3.4)
Senses[2]	8.1 (4.3)	8.0 (3.3)	8.0 (4.7)	8.0 (4.4)
Imageability	4.8 (1.2)	4.9 (0.9)	5.1 (0.8)	4.8 (1.1)
Concreteness	5.3 (0.8)	5.2 (0.7)	5.0 (0.8)	5.0 (0.9)
Age of acquisition	7.6 (1.8)	7.0 (2.0)	7.0 (1.6)	6.9 (1.8)
Semantic diversity	1.6 (0.2)	1.6 (0.2)	1.6 (0.2)	1.7 (0.2)
Subjective familiarity	5.0 (0.4)	5.0 (0.6)	5.0 (0.5)	5.1 (0.5)

Note: Standard deviations are given in the parentheses. Each prime set contained 28 items. Meaning-relatedness and dominance ratings come from Experiment 1. Word-form frequency, bigram frequency, and the number of orthographic neighbours come from the BNC (2007). "Senses[1]" and "Senses[2]" refer to the number of senses in the Wordsmyth (Parks et al., 1998) and the WordNet Dictionary (Fellbaum, 1998), respectively. Imageability, concreteness, and familiarity ratings come from the MRC Psycholinguistic Database (Coltheart, 1981). Age-of-acquisition ratings come from Kuperman et al. (2012). Semantic diversity data come from Hoffman et al. (2013).

The four sets of primes were statistically comparable at the group level (all *F*s < 1) with respect to 14 lexical and semantic variables, such as word-form frequency and imageability (see stimulus properties in Table 3.1 above). None of the primes was a compound word (e.g., "childhood") or a homophone[13]. Critically, the sets were matched on the number of word senses to ensure that the

[13] Homophones are words that have identical pronunciation but different spelling and meanings (e.g., "maid" vs. "made"). Homophones are not suitable for the present line of research because they may produce additional competition at the phonological level.

experimental manipulation reflects the impact of homonymy, rather than that of polysemy. Prime-word selection was also constrained by word length (3-6 letters) and word-form frequency (5-60 occurrences per million) in the BNC (2007). Due to a small number of balanced homonyms in English (Armstrong et al., 2012), it was impossible to test stimuli of a single syntactic class. This variability was, however, controlled across the four sets of primes. Most words in each set had either noun-noun (range across the sets: 12-17) or noun-verb interpretations (7-9), with very few having noun-adjective (1-4), verb-adjective (0-3), adjective-adjective (0-1), or verb-verb interpretations (0-1).

Each of the homonyms was paired with two semantically related targets. For unbalanced homonyms, the first target ("HF-meaning") related to the dominant meaning of the prime (e.g., "pen-ink"), while the other ("LF-meaning") related to the alternative meaning (e.g., "pen-farmer"). The same split was used for balanced homonyms (e.g., "fan-cheer" vs. "fan-breeze"), although the difference in meaning frequencies for these items was considerably smaller, as evident in the norms from Experiment 1. None of the targets was related to both meanings of the ambiguous word. Non-homonyms were paired with two targets (A & B) that related to the same interpretation/sense of the word (e.g., "fake-truth" vs. "fake-fraud"). This aimed to equalise prime repetition and the number of items in the contrast between ambiguous and unambiguous words. Most of the prime-target pairs were related through action-recipient relationship (e.g., "jam-knife"), category membership (e.g., "novel-poem"), physical properties (e.g., "temple-chapel"), and synonymy (e.g., "lean-slim"). Stimulus selection excluded target words that were lexically, rather than semantically, associated with prime words (e.g., "tap-water"). All primes were also paired with two targets (A & B) that were unrelated to either of their meanings (e.g., “fan-snake” vs. “fan-cancel”). The number of unrelated targets was doubled in order to equalise the number of related and unrelated trials in the experiment. All target words had only one meaning in the Wordsmyth Dictionary (Parks et al., 1998), except for “chess” which had two additional LF meanings (a floor board of a pontoon bridge; a variety of weedy grasses). All 16 sets of prime-target pairs were matched (all *F*s < 1) on 14 lexical and semantic variables, such as word-form frequency and age of acquisition (see Tables 3.2 & 3.3 below). The syntactic class of the targets was balanced across the sets, such that each contained 17-23 nouns, 3-7 verbs, and 2-6 adjectives. Overall, the stimulus set comprised 224 related and 224 unrelated word pairs (see Appendix 2).

Table 3.2: Experiment 2: Properties of the related word pairs.

Variable	Homonymous prime				Non-homonymous prime			
	Balanced		Unbalanced		Set 1		Set 2	
	HF	LF	HF	LF	A	B	A	B
Example	"fan-cheer"	"fan-breeze"	"pen-ink"	"pen-farmer"	"fake-truth"	"fake-fraud"	"fur-fox"	"fur-rabbit"
Prime-target relatedness	5.8 (0.6)	5.8 (0.4)	6.01 (0.5)	4.4 (1.1)	6.0 (0.5)	5.9 (0.5)	5.9 (0.5)	6.0 (0.5)
Letters	5.0 (1.0)	5.1 (1.1)	4.7 (1.1)	5.1 (1.2)	4.9 (1.0)	4.6 (0.9)	4.7 (0.9)	4.8 (0.8)
Phonemes	4.1 (0.9)	4.3 (1.3)	3.9 (1.0)	4.1 (1.4)	4.0 (0.8)	3.8 (0.9)	3.9 (0.9)	3.9 (0.8)
Syllables	1.3 (0.5)	1.6 (0.7)	1.4 (0.6)	1.5 (0.6)	1.4 (0.6)	1.2 (0.4)	1.3 (0.5)	1.4 (06)
Raw frequency	24.4 (14.7)	24.9 (19.9)	24.2 (18.1)	28.1 (25.4)	28.0 (24.4)	26.1 (22.9)	24.4 (20.6)	27.1 (23.4)
Log frequency	1.3 (0.3)	1.3 (0.4)	1.3 (0.3)	1.3 (0.5)	1.3 (0.4)	1.2 (0.4)	1.2 (0.4)	1.3 (0.4)
Orthographic neighbours	4.1 (4.8(	3.8 (5.0)	5.2 (3.6)	4.3 (5.3)	4.1 (5.1)	5.1 (4.8)	5.1 (4.6)	5.3 (5.6)
Log bigram frequency	2.7 (0.5)	2.6 (0.5)	2.8 (0.6)	2.8 (0.4)	2.8 (0.3)	2.8 (0.4)	2.8 (0.4)	2.8 (0.4)
Senses[1]	5.5 (3.4)	5.1 (3.2)	5.3 (3.3)	5.3 (2.2)	5.5 (3.9)	5.4 (3.4)	5.4 (5.9)	5.9 (4.2)
Senses[2]	6.1 (5.0)	5.3 (3.9)	5.7 (3.1)	5.6 (3.7)	6.0 (3.8)	5.4 (3.0)	5.6 (3.5)	6.1 (4.0)
Imageability	5.2 (1.1)	4.8 (1.3)	5.1 (1.0)	5.6 (1.0)	4.9 (1.0)	5.2 (1.1)	5.2 (0.8)	5.1 (1.0)
Concreteness	5.5 (0.9)	5.1 (1.0)	5.1 (0.9)	5.6 (0.9)	5.0 (1.1)	5.1 (1.0)	4.9 (1.2)	5.2 (0.9)
Age of acquisition	6.2 (1.6)	7.1 (2.1)	6.2 (1.5)	5.5 (2.4)	6.6 (1.9)	6.3 (1.8)	6.3 (2.1)	6.2 (2.5)
Semantic diversity	1.6 (0.2)	1.6 (0.3)	1.6 (0.2)	1.6 (0.2)	1.7 (0.3)	1.7 (0.2)	1.6 (0.3)	1.6 (0.2)
Subjective familiarity	5.4 (0.5)	5.3 (0.6)	5.4 (0.6)	5.4 (0.5)	5.3 (0.4)	5.2 (0.5)	5.2 (0.6)	5.3 (0.5)

Note. Standard deviations are given in the parentheses. Information on the different variables is given in the note for Table 3.1.

Table 3.3: Experiment 2: Properties of the unrelated word pairs.

Variable	Homonymous prime				Unambiguous prime			
	Balanced		Unbalanced		Set 1		Set 2	
	HF	LF	HF	LF	HF	LF	HF	LF
Example	“fan-snake”	“fan-cancel”	“pen-yeast”	“pen-add“	“fake-expand”	“fake-fetch”	“fur-chain”	“fur-pill”
Prime-target relatedness	2.1 (0.6)	2.1 (0.5)	1.9 (0.3)	1.9 (0.3)	1.9 (0.3)	2.0 (0.4)	1.9 (0.4)	2.0 (0.4)
Letters	4.7 (0.9)	4.7 (0.8)	4.7 (1.0)	4.6 (1.1)	4.7 (0.8)	4.6 (0.8)	4.7 (1.1)	4.7 (1.0)
Phonemes	4.0 (0.8)	4.0 (1.2)	3.9 (1.0)	3.8 (1.0)	3.8 (1.0)	3.8 (1.0)	4.0 (1.1)	4.0 (1.0)
Syllables	1.4 (0.5)	1.4 (0.5)	1.3 (0.5)	1.3 (0.5)	1.4 (0.5)	1.3 (0.5)	1.4 (0.6)	1.4 (0.5)
Raw frequency	26.1 (23.7)	25.5 (22.8)	25.5 (19.0)	25.9 (22.5)	24.4 (23.1)	25.5 (23.6)	25.2 (22.4)	24.7 (24.2)
Log frequency	1.2 (0.4)	1.2 (0.4)	1.3 (0.4)	1.3 (0.4)	1.2 (0.2)	1.2 (0.4)	1.2 (0.4)	1.2 (0.5)
Orthographic neighbours	5.2 (5.1)	5.0 (5.6)	5.3 (5.4)	5.1 (4.4)	5.6 (5.2)	5.3 (5.2)	5.2 (6.0)	5.4 (5.6)
Log bigram frequency	2.7 (0.4)	2.7 (0.5)	2.7 (0.4)	2.6 (0.5)	2.7 (0.4)	2.6 (0.5)	2.6 (0.5)	2.6 (0.4)
Senses[1]	5.1 (2.6)	5.1 (3.2)	5.3 (2.7)	5.1 (2.7)	5.4 (3.1)	5.1 (1.8)	5.1 (3.2)	5.1 (2.4)
Senses[2]	5.7 (2.9)	5.7 (3.5)	5.6 (3.5)	5.7 (4.6)	5.7 (3.5)	5.6 (3.7)	5.5 (3.5)	5.7 (3.7)
Imageability	5.2 (1.0)	5.3 (1.0)	5.2 (0.9)	5.1 (1.0)	5.2 (1.1)	5.3 (0.8)	5.2 (0.9)	5.2 (1.1)
Concreteness	5.2 (0.8)	5.1 (1.0)	5.2 (1.1)	5.3 (1.0)	5.0 (1.1)	5.1 (0.9)	4.9 (1.1)	5.3 (1.1)
Age of acquisition	6.2 (1.9)	6.4 (2.3)	6.1 (2.1)	6.0 (1.8)	6.1 (1.7)	6.2 (2.1)	6.1 (1.7)	6.2 (2.1)
Semantic diversity	1.6 (0.2)	1.6 (0.2)	1.5 (0.3)	1.6 (0.2)	1.6 (0.2)	1.6 (0.3)	1.6 (0.2)	1.6 (0.2)
Subjective familiarity	5.2 (0.6)	5.3 (0.7)	5.3 (0.6)	5.3 (0.5)	5.3 (0.7)	5.2 (0.8)	5.3 (0.5)	5.2 (0.5)

Note: Standard deviations are given in the parentheses. Information on the different variables is given in the note for Table 3.1.

The semantic relatedness between the prime and the target word was confirmed by a group of 40 monolingual British-English native speakers [17 females, aged 18-40 (*M* = 28.3, *SD* = 6.4)]. Participants rated all word pairs on a 7-point scale (where 1 denoted "highly unrelated" and 7 "highly related"). For unrelated pairs, the maximum prime-target relatedness rating was set to 3.3 (corresponding to "slightly unrelated"). For related pairs, the minimum rating was 4.5 (between "neither related nor unrelated" and "slightly related"). Eleven unrelated and 51 related targets that did not meet these criteria were replaced by new targets that were then rated, using the same procedure, by another 40 monolingual native speakers [23 females, aged 19-38 (*M* = 28.8, *SD* = 6.1)]. Mean relatedness ratings for the final sets of related and unrelated pairs are given in Tables 3.2 and 3.3, respectively. There were no significant differences among the final sets of unrelated pairs for all word types [$F(7, 216) = 0.9$, $p = .53$). However, for the related counterparts, ratings were significantly lower (all *p*s < .001) for the LF-meaning targets of unbalanced homonyms than for every other, otherwise well-matched, set.

Twenty monolingual British-English native speakers [8 females, aged 21-36 (*M* = 29.3, *SD* = 5.3)] participated in another prime-target relatedness pre-test. The aim of this study was to establish whether the low ratings of the relatedness between unbalanced homonyms and their targets referring to the LF meaning truly reflected poor semantic relatedness, or whether participants' ratings were biased by meaning dominance. In this pre-test, participants read a short sentence (all taken from the Oxford Dictionary; Simpson & Weiner, 1989) containing a homonym in its LF-meaning context (e.g., "Along with the original small red house, we now have two barns, a sheep <u>pen</u> and several sheds") and then rated the relatedness between the homonym and four words – namely, two unrelated fillers and two targets related to the LF meaning. Critically, these ratings were collected for the LF-meaning targets of balanced and unbalanced homonyms. The “primed” ratings (all at or above 4.5) for the LF-meaning targets of unbalanced homonyms were significantly higher (M = 5.6, SD = 0.5) than those in the previous “unprimed” pre-test [M = 4.4, SD = 1.1, $t(27) = 6.6$, $p < .001$] and no longer significantly lower than the ratings for the other sets of related pairs [$F(7, 216) = 1.8$, $p = .09$]. Interestingly, the primed ratings for the LF-meaning targets of balanced homonyms (M = 5.7, SD = 0.5) did not differ from those made in the absence of context [M = 5.8, SD = 0.4; $t(27) = 0.6$, $p = .58$]. Taken together, these results suggest that the LF-meaning targets of unbalanced homonyms were indeed related and suitable for testing. The pairs were given relatively lower relatedness ratings in the absence of context only because of participants’ bias towards the dominant meaning, and not because they were not related to the homonym.

3.2.1.3 Procedure

Participants decided whether the prime and the target word were related in meaning (for the rationale behind the task, see 1.5 Aims of the Thesis in Chapter 1) by pressing keyboard buttons (L for “yes” with their dominant, right hand, A for “no” with their left hand). The task was programmed in EPrime 2.0 (Schneider, Eschman, & Zuccolotto, 2012). Stimuli were pseudo-randomly divided into four blocks of 112 trials, such that each block contained the same number of related and unrelated word pairs, the same number of the four primes, and the same number of the two targets. Participants responded four times to each prime and once to each target. None of the primes appeared more than once within the same block. The order of the blocks was counter-balanced across participants. The order of the trials within each block was pseudo-randomised, such that no more than three related/unrelated word pairs appeared consecutively. The task began with a 32-trial practice block that comprised examples of each condition. In this practice phase, participants received feedback on both the speed and accuracy of their responses. There were two one-minute breaks - one after the practice block and one 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. Trials began with a 500 ms fixation cross. After a delay of 100 ms, the prime and the target were presented for 200 ms and 500 ms, respectively, with a 50-ms delay in between (for an illustration of the trial procedure, see Figure 3.1 below). Once the target disappeared, there was a 1500 ms delay for response execution followed by a 100 ms inter-trial interval (ITI). 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 to slow and would be excluded from analyses). Response speed and accuracy were equally emphasised in the instructions, and participants were told what constituted semantic relatedness and given examples. On average, testing lasted for 24 minutes.

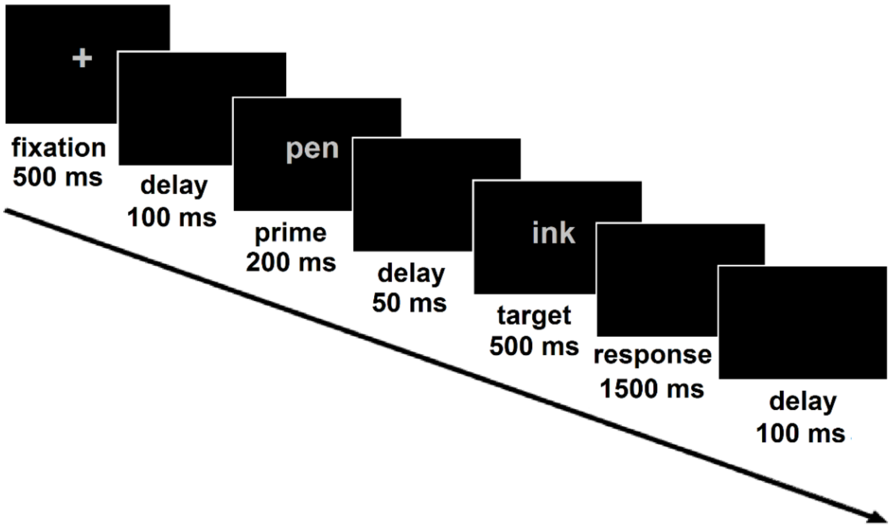

Figure 3.1: Experiment 2: Illustration of the trial sequence.

3.2.2 Results

Two of the 35 participants were removed from all analyses – one due to a large number of errors on related trials (63.8%) and the other due to slow and variable performance across all trials (*M* = 899.9 ms, *SD* = 182.9). The four targets of the non-homonym “elbow” were removed as they were inadvertently paired with a different prime during the experiment. For response times (RTs), analyses also excluded incorrect responses (19.2% of all trials) and outliers (two standard deviations above/below a participant’s mean per condition; 4.1% of all trials). The remaining RTs were log-transformed to further minimise the impact of potential outliers and to normalise residual distributions.

Although the ambiguity disadvantage effect manifests itself in the speed rather than the outcome of disambiguation, analyses included both accuracy and latency data given that findings from accuracy analyses provide invaluable help when interpreting findings from latency analyses. Accuracy and latency data were analysed 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)[14]. All response-latency models additionally included the factor of Block (1, 2, 3, 4). Main effects and interactions involving this factor are not reported as its sole purpose was to account for potential practice or prime-repetition effects (Pollatsek & Well, 1995), and no such effects were found in any of the experiments[15]. Terms involving Block were removed from response-accuracy analyses because they resulted in model non-convergence. Responses to related and unrelated word pairs were modelled separately, as analyses involving both types of trials showed a significant effect of Trial (i.e., faster but less accurate responses on related trials) that always interacted with the effects of Prime and Target.

Each model included significant random intercepts for subjects and items. Following Barr et al. (2013) and Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017), the optimal random-effects structure justified by the data was identified using the forward model selection method[16]. The only random slope that

---

[14] Throughout this thesis, data from lab-based experiments was analysed using mixed-effects modelling. This approach captures and accounts for variability in effects across individuals and items (see Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013), maximising detection power without increasing Type 1 error. Mixed-effects modelling has also been shown to be insensitive to unbalanced designs (see Pinheiro & Bates, 2000), and as such it is highly suitable for the current experiments with unequal numbers of correct responses per cell.

[15] Analyses for Experiments 2-4 revealed that performance did not reliably change across the blocks of these short experiments. Likewise, there was no indication that having responded to an ambiguous word in one meaning affected the processing of that word on subsequent trials instantiating the other meaning, as a result of strategic processing or word-meaning priming.

[16] In this procedure, analysis begins with a model that includes significant random intercepts and tests all possible slopes for inclusion separately. Out of significant slopes, researchers first add the most influential one (based on the value of χ^2 from model-comparison tests) to the base model and then test whether the second most influential slope further improves the model. The remaining slopes are tested and added until the model fails to converge.

significantly improved fit and was included in response-latency 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 prevent spurious results. Following Nakagawa and Schielzeth (2013) and Johnson (2014), marginal R^2 (variance explained only by fixed effects) and conditional R^2 (variance explained by both fixed and random effects) for all models were estimated using the “MuMIn” package (Bartoń, 2014). For RT results, back-transformed means and their 95% confidence intervals (CIs) that were estimated from the mixed-effects models using the “lmerTest” package (Kuznetsova, Brockhoff, & Christensen, 2017) are reported below.

3.2.2.1 Related Trials

Mean error rates (%) and RTs (ms) across related word pairs are illustrated in Figure 3.2 below. The response-accuracy model for related trials (marginal R^2 = .18, conditional R^2 = .45) revealed a significant effect of Prime [$\chi^2(3) = 63.1$, $p < .001$]; error rates were higher for both balanced (*M* = 39.2%, *SD* = 11.7; $p < .001$) and unbalanced homonyms (*M* = 44.1%, *SD* = 11.8; $p < .001$) than non-homonyms (*M* = 18.4%, *SD* = 10.8). Relatedness decisions were generally more erroneous to LF-meaning (*M* = 39.5%, *SD* = 10.6) than HF-meaning targets (*M* = 20.7%, *SD* = 11.0), and this effect of Target [$\chi^2(1) = 37.9$, $p < .001$] interacted with the type of Prime [$\chi^2(3) = 39.7$, $p < .001$]. Relative to both targets of non-homonyms (M_A = 16.7%, *SD* = 11.2; M_B = 20.1%, *SD* = 11.6), error rates were significantly higher for the HF-meaning (*M* = 29.3%, *SD* = 12.4; $p < .01$) and LF-meaning targets of balanced homonyms (*M* = 49.0%, *SD* = 13.5; $p < .001$), the LF-meaning targets of unbalanced homonyms (*M* = 68.2%, *SD* = 14.0; $p < .001$), but not the HF-meaning targets of unbalanced homonyms (*M* = 20.0%, *SD* = 13.7; $p = .33$).

The latency model (marginal $R^2 = .04$, conditional $R^2 = .51$) revealed a significant effect of Prime [$\chi^2(3) = 39.1$, $p < .001$]; RTs were higher for both balanced (*M* = 632.0 ms, 95% CIs: 599.5, 666.4; $p < .001$) and unbalanced homonyms (*M* = 640.6 ms, 95% CIs: 607.3, 675.6; $p < .001$) than the non-homonymous counterparts (*M* = 594.5 ms, 95% CIs: 564.1, 626.7). Responses were generally slower to LF-meaning (*M* = 628.4 ms, 95% CIs: 596.9, 661.5) than HF-meaning targets (*M* = 602.0 ms, 95% CIs: 572.0, 633.6), and this effect of Target [$\chi^2(1) = 13.7$, $p < .001$] interacted with the type of Prime [$\chi^2(3) = 26.2$, $p < .001$]. Relative to both targets of non-homonyms (M_A = 594.9 ms, 95% CIs: 562.8, 628.8; M_B = 594.2 ms, 95% CIs: 562.1, 628.2), relatedness decisions were slower to the LF-meaning targets of balanced (*M* = 647.3 ms, 95% CIs: 611.7, 685.0; $p < .001$) and unbalanced homonyms (*M* = 682.3 ms, 95% CIs: 643.7, 723.1; $p < .001$), the HF-meaning targets of balanced homonyms (*M* = 617.2 ms, 95% CIs: 583.7, 652.5; $p < .05$), but not the HF-meaning targets of unbalanced homonyms (*M* = 601.5 ms, 95% CIs: 569.0, 635.8; $p = .49$).

Further analysis of RTs on related trials included properties of the prime and the target as covariates. The rationale was that due to a large number of errors for homonyms in the LF meaning, these conditions involved subsets of items that were no longer matched with their non-homonymous counterparts. Each of the variables used for group-level matching of prime (see Table 3.1 above) and target words (see Table 3.2 above) was considered for inclusion. However, to prevent model over-fitting, analysis included only those variables that significantly correlated with RT – namely, age of acquisition, number of orthographic neighbours, number of related senses, imageability[17], and syntactic ambiguity (i.e., whether the different meanings of words corresponded to the same parts of speech) at the prime level as well as prime-target relatedness, number of orthographic neighbours, imageability, subjective familiarity, and age of acquisition at the target level. The effects of Prime and Target reported above were still significant when these variables were taken into account, suggesting that the results were due to the homonymous status of the words that the experiment explicitly manipulated, rather than due to unsystematic differences in the controlled properties of the items (e.g., word-form frequency, imageability, age of acquisition).

[17] Imageability was chosen over concreteness due to the greater availability of item ratings in the MRC Psycholinguistic Database (Coltheart, 1981).

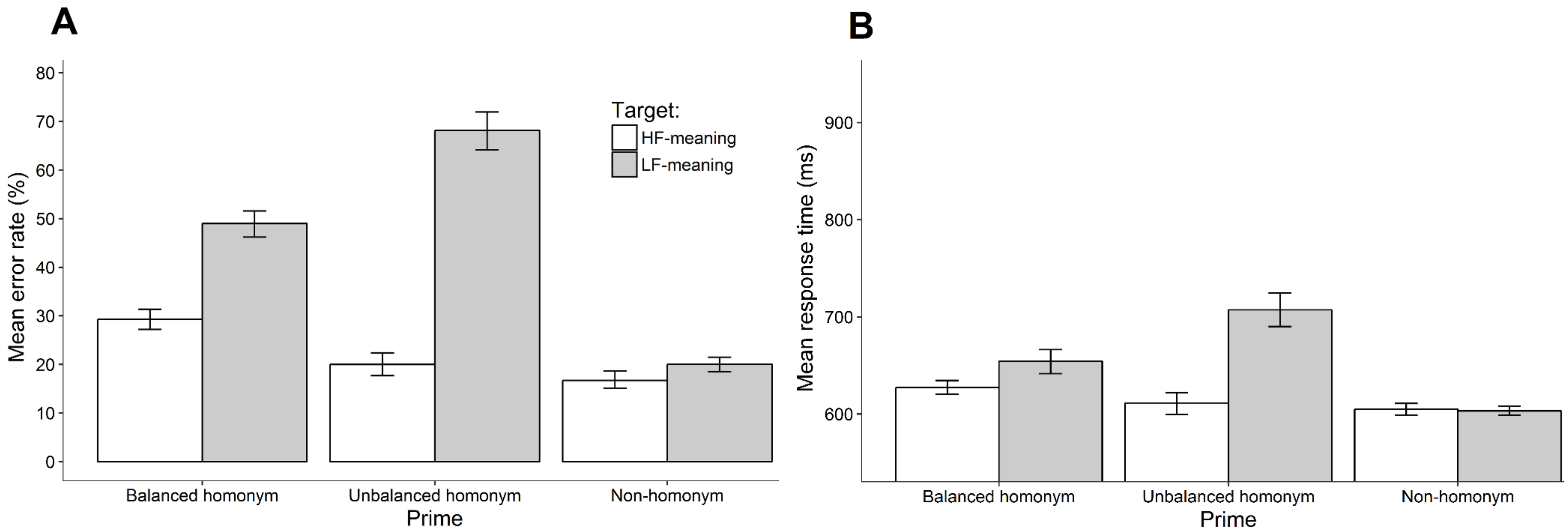

Figure 3.2: Experiment 2: Subject means of error rates (Panel A) and untransformed RTs (Panel B) across related word pairs. Error bars show 95 % confidence intervals adjusted to remove between-subjects variance (Loftus & Masson, 1994).

3.2.2.2 Unrelated Trials

Mean error rates and RTs across unrelated word pairs are illustrated in Figure 3.3 below. The response-accuracy model for unrelated trials (marginal R^2 = .01, conditional R^2 = .56) revealed a significant effect of Prime [$\chi^2(3) = 10.3$, $p < .05$]. Compared to non-homonyms (*M* = 8.2%, *SD* = 6.3), error rates were higher for balanced homonyms (*M* = 10.3%, *SD* = 9.0) but lower for unbalanced homonyms (*M* = 6.1%, *SD* = 7.1), though neither contrast was significant after the Holm-Bonferroni correction (both *p*s = .10).

In the response-latency model (marginal R^2 = .01, conditional R^2 = .56) there was a significant effect of Prime [$\chi^2(3) = 12.8$, $p < .01$]. Compared to non-homonyms (*M* = 626.3 ms, 95% CIs: 592.1, 662.7), RTs were significantly slower for balanced (*M* = 643.3 ms, 95% CIs: 608.0, 680.5; $p < .01$) but not unbalanced homonyms (*M* = 633.6 ms, 95% CIs: 599.0, 670.3; $p = .16$). All other effects were non-significant.

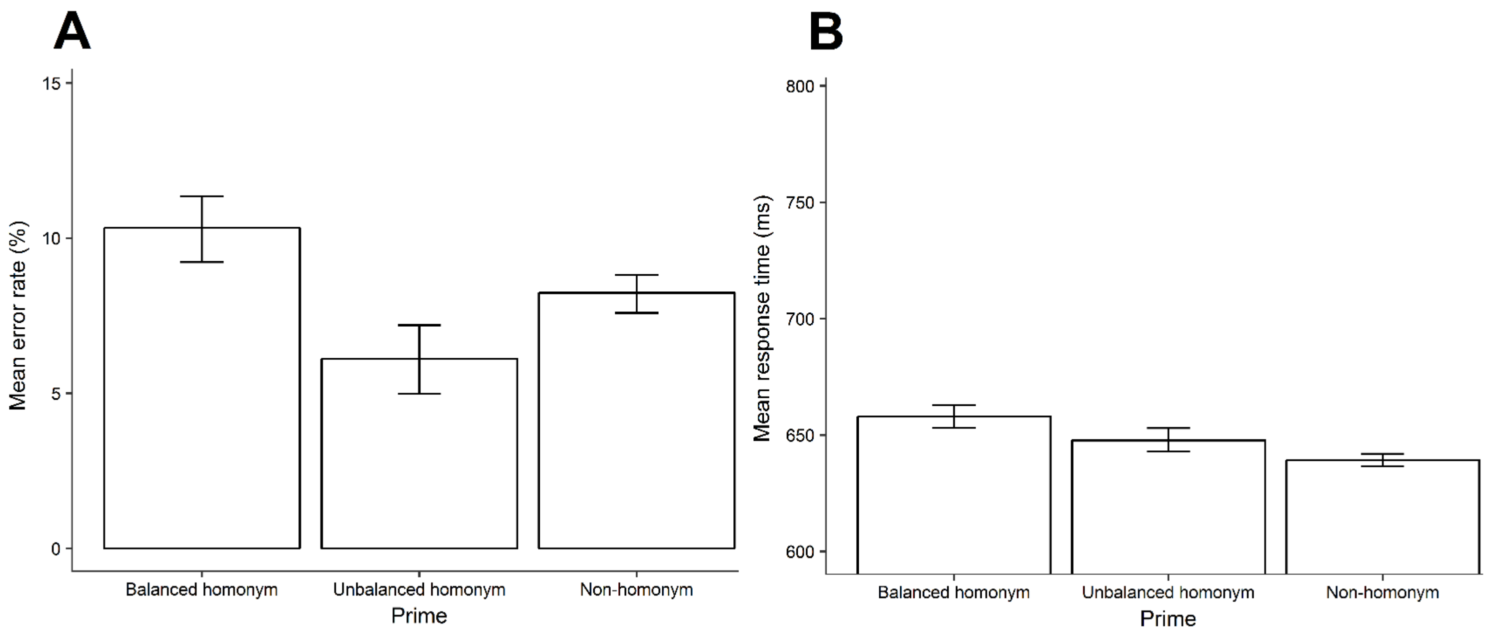

Figure 3.3: Experiment 2: Subject means of error rates (Panel A) and untransformed RTs (Panel B) across unrelated word pairs. Error bars show 95 % confidence intervals adjusted to remove between-subjects variance.

3.2.3 Discussion

Experiment 2 showed that meaning frequency modulates the ambiguity disadvantage. For unbalanced homonyms, there was a substantial processing cost when these words were disambiguated toward the LF but not the HF meaning. Together with the null effect of ambiguity on unrelated trials for these homonyms, the results suggest that the meanings of unbalanced homonyms do not compete for semantic activation, most likely due to weak and delayed activation of the alternative meaning (Frost & Bentin, 1992; Simpson & Burgess, 1985). Unbalanced homonymy slows task performance 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.

The pattern of responses was different for balanced homonyms. These words incurred a processing cost regardless of which meaning was implied. Critically, the processing cost also appeared for responses to unrelated pairs, providing initial evidence against the decision-making account (Pexman et al., 2004), according to which the ambiguity effect should only arise on related trials when the meanings of the ambiguous prime trigger conflicting responses to the target. Overall, consistent with Rayner and Duffy (1986) and Duffy et al. (1988), Experiment 2 suggests that the disadvantage effect is restricted to homonyms with more balanced meaning frequencies. It also shows, for the first time, that the effect may indeed lie in semantic competition, rather than response conflict, as it also arose on unrelated trials that do not involve such conflict.

3.3 Experiment 3: Long Prime Duration

The leading aim of Experiment 3 was to investigate the impact of prime-word duration on the comprehension of homonymy. In particular, Experiment 3 explored the possibility that unbalanced homonyms in Experiment 2 were generally processed as fast as their unambiguous counterparts because the task did not provide enough time to retrieve their LF meaning, thus reducing potential semantic competition.

This explanation is, however, unlikely. First, the prime duration of 200 ms in Experiment 2 mirrors readers' fixation durations (260 ms) for unbalanced homonyms in eye-movement studies (Duffy et al., 1988; Rayner & Duffy, 1986), which suggests that the experiment provided participants with approximately as much time to process the words as they would require when reading them under more natural circumstances. Second, research into meaning activation (Frost & Bentin, 1992; Simpson & Burgess, 1985) has demonstrated that 100-300 ms is sufficient for the LF meaning to reach its maximal, albeit still relatively weak, activation. However, since this research did not specifically examine the patterns of meaning activation for unbalanced homonyms, Experiment 3 was necessary to confirm that these words do not show a disadvantage effect, even when readers have enough time to retrieve the alternative meaning. To address this issue, Experiment 3 involved the same task and stimuli as Experiment 2, but the primes were presented for a longer period of time (700 ms).

The second aim of Experiment 3 was to investigate the cause of the relatively high proportion of errors on related trials in Experiment 2. While errors were expected to be more common for homonyms in the LF- than the HF-meaning condition (Bitan, Kaftory, Meiri-Leib, Eviatar, & Peleg, 2017; Harpaz, Lavidor, & Goldstein, 2013; Pexman et al., 2004), participants' performance was surprisingly error-prone across all sets of related prime-target word pairs, even those involving non-homonyms.

One possibility is that the fast-paced nature of the task in Experiment 2 compromised accuracy. Unlike previous relatedness decision studies (Gottlob et al., 1999; Pexman et al., 2004; Piercey & Joordens, 2000), in which the prime was presented for 1000 ms or simultaneously with the target and remained visible until a response was made, the prime and the target in Experiment 2 were presented only for 200 ms and 500 ms, respectively. Although the aim was to promote fast performance unaltered by decision making and strategic stimulus processing (see Hill, Strube, Roesch-Ely, & Weisbrod, 2002; Neely, 1977; Rossell, Bullmore, Williams & David, 2001), it is plausible that the short stimulus presentation produced a speed-accuracy trade-off. The prime duration of 700 ms in Experiment 3 should provide participants with more time to detect and judge the semantic relatedness between the prime and the target, and thus establish whether the high proportion of errors on related trials in Experiment 2 was due to the speeded nature of the task.

3.3.1 Method

3.3.1.1 Participants

A different group of 30 students or members of staff [21 females, aged 18-42 (*M* = 21.3, *SD* = 5.5)] from the University of Leeds participated in the experiment in exchange for four course credits or £3. Participants were right-handed monolingual native speakers of British English with no known history of language-vision-related difficulties or disorders. The experiment received ethical approval from the School of Psychology, University of Leeds Ethics Committee.

3.3.1.2 Stimuli & Procedure

The same stimuli and procedure as in Experiment 2 were used. The only difference between the two experiments was the longer prime-word duration in Experiment 3 (700 ms instead of 200 ms). On average, testing lasted for 28 minutes.

3.3.2 Results

One of the 30 participants was removed from all analyses due to a large number of errors across all related trials (54.5%). For RTs, analyses excluded errors (17.2% of all trials) and outliers (two standard deviations above/below a participant's mean per condition; 3.4% of all trials). Both accuracy and RT data were analysed in the same way as in Experiment 2.

3.3.2.1 Related Trials

Mean error rates and RTs across related word pairs are illustrated in Figure 3.4. The response-accuracy model for related trials (marginal R^2 = .26, conditional R^2 = .52) revealed a significant effect of Prime [$\chi^2(3) = 90.5$, $p < .001$];

error rates were higher for both balanced (*M* = 40.5%, *SD* = 12.4; $p < .001$) and unbalanced homonyms (*M* = 50.1%, *SD* = 11.2; $p < .001$) than non-homonyms (*M* = 15.5%, *SD* = 7.6). Responses were generally more erroneous to LF-meaning (*M* = 41.1%, *SD* = 10.3) than HF-meaning targets (*M* = 19.7%, *SD* = 8.3), and this effect of Target [$\chi^2(1) = 41.2$, $p < .001$] interacted with the type of Prime [$\chi^2(3) = 46.4$, $p < .001$]. For LF-meaning targets, error rates were higher for balanced (*M* = 51.5%, *SD* = 16.5; $p < .001$) and unbalanced homonyms (*M* = 78.2%, *SD* = 15.7; $p < .001$) than non-homonyms (*M* = 17.4%, *SD* = 7.6). For HF-meaning targets, error rates were also higher for balanced (*M* = 29.4%, *SD* = 11.2; $p < .001$) and unbalanced homonyms (*M* = 22.0%, *SD* = 9.7; $p < .05$) than non-homonyms (*M* = 13.7%, *SD* = 8.5), but to a significantly smaller extent.

The response-latency model (marginal $R^2 = .05$, conditional $R^2 = .44$) revealed a significant effect of Prime [$\chi^2(3) = 36.9$, $p < .001$]; responses were slower to balanced (*M* = 783.4 ms, 95% CIs: 744.7, 824.1; $p < .001$) and unbalanced homonyms (*M* = 792.9 ms, 95% CIs: 752.3, 835.4; $p < .001$) than non-homonyms (*M* = 718.9 ms, 95% CIs: 683.4, 756.3). RTs were generally higher for LF-meaning (*M* = 782.3 ms, 95% CIs: 744.9, 821.5) than HF-meaning targets (*M* = 721.4 ms, 95% CIs: 687.4, 757.4), and this effect of Target [$\chi^2(1) = 22.0$, $p < .001$] interacted with the type of Prime [$\chi^2(3) = 39.0$, $p < .001$]. Relative to both targets of non-homonyms (M_A = 717.2 ms, 95% CIs: 678.8, 757.8; M_B = 720.7 ms, 95% CIs: 682.0, 761.6), responses were slower to the HF- (*M* = 752.8 ms, 95% CIs: 712.5, 795.4; $p < .05$) and LF-meaning targets of balanced homonyms (*M* = 815.8 ms, 95% CIs: 770.9, 863.6 ms; $p < .001$), the LF-meaning targets of unbalanced homonyms (*M* = 884.1 ms, 95% CIs: 832.0, 939.7; *p* <

.001), but not the HF-meaning targets of unbalanced homonyms (*M* = 710.9 ms, 95% CIs: 672.7, 751.3; $p = .65$).

As in Experiment 2, additional analyses explored whether the effects of Prime and Target on RTs were still significant after the lexical and semantic properties of the words had been taken into account. Following the same procedures for covariate selection, the analyses included the number of senses and syntactic ambiguity at the prime level as well as prime-target relatedness and imageability at the target level. The results remained significant, except for the effect for balanced homonyms in the HF meaning that only approached the significance threshold ($p = .07$) after the Holm-Bonferroni adjustment.

3.3.2.2 Unrelated Trials

Mean error rates and RTs across unrelated word pairs are illustrated in Figure 3.5. The response-accuracy model for unrelated trials (marginal $R^2 = .02$, conditional $R^2 = .25$) revealed an unexpected marginal interaction between the effects of Prime and Target [$\chi^2(3) = 7.7$, $p = .05$] that was due to numerically higher error rates for the LF-meaning targets of balanced homonyms (*M* = 6.2%, *SD* = 6.1) than one of the two sets of targets paired with non-homonyms (*M* = 3.8%, *SD* = 4.3). This contrast was not, however, significant ($p = .14$) after the Holm-Bonferroni correction.

As in Experiment 2, in the response-latency model (marginal $R^2 = .01$, conditional $R^2 = .48$) there was a significant effect of Prime [$\chi^2(3) = 16.9$, $p < .001$]. Compared to non-homonyms (*M* = 726.2 ms, 95% CIs: 688.3, 766.0), responses were slower to balanced (*M* = 751.5 ms, 95% CIs: 712.4, 792.9; *p* <

.001) but not unbalanced homonyms (*M* = 736.2 ms, 95% CIs: 697.9, 776.6; *p* = .10). All other effects were non-significant.

3.3.2.3 Experiments 2 & 3 Combined: Related Trials

A key question raised in the introduction to Experiment 3 was whether longer prime duration would provide more time to retrieve the LF meanings of unbalanced homonyms, such that they would interfere with the HF meanings. In order to explore the impact of prime duration on the processing of homonymy, supplementary analyses contrasted task performance in Experiments 2 and 3. These analyses were the same as those conducted for each experiment separately, except that they included the additional factor of Experiment. All models included significant random intercepts for subjects and items as well as a random slope for Experiment across subjects. The random by-subjects slope for Block in response-latency models was removed due to non-convergence. Planned contrasts (with the Holm-Bonferroni correction) explored between-experiments differences in the previously reported effects of Prime and Target.

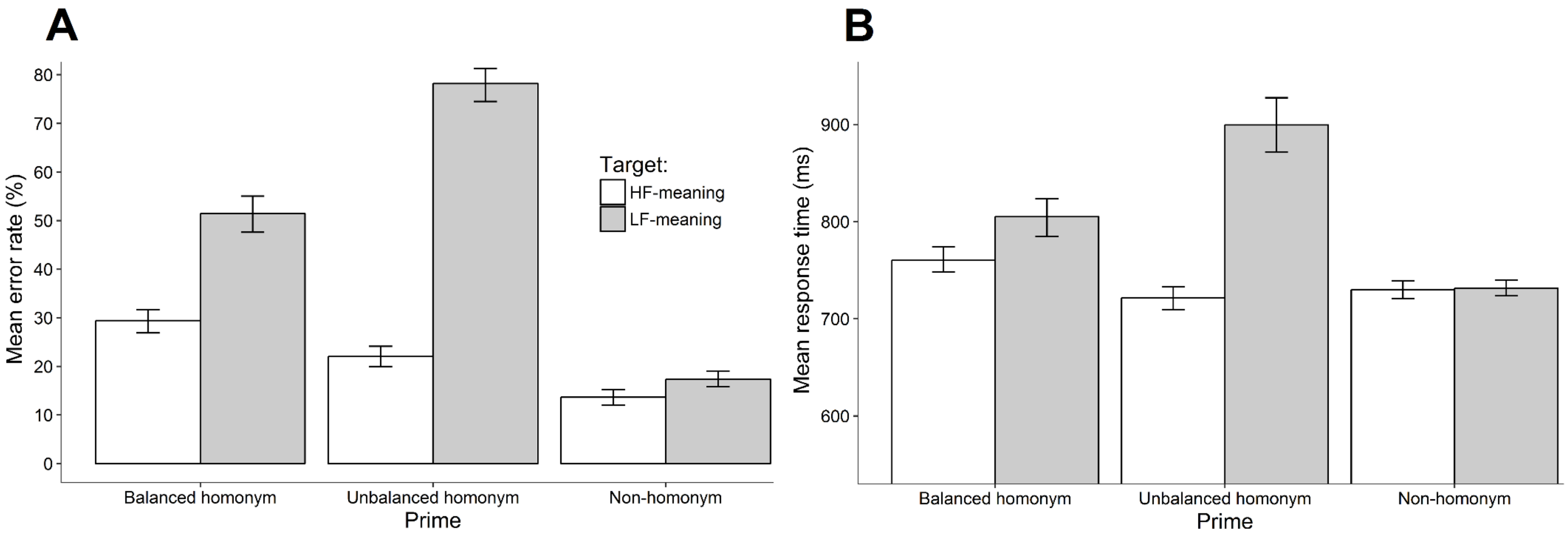

Figure 3.4: Experiment 3: Subject means of error rates (Panel A) and untransformed RTs (Panel B) across related word pairs. Error bars show 95 % confidence intervals adjusted to remove between-subjects variance.

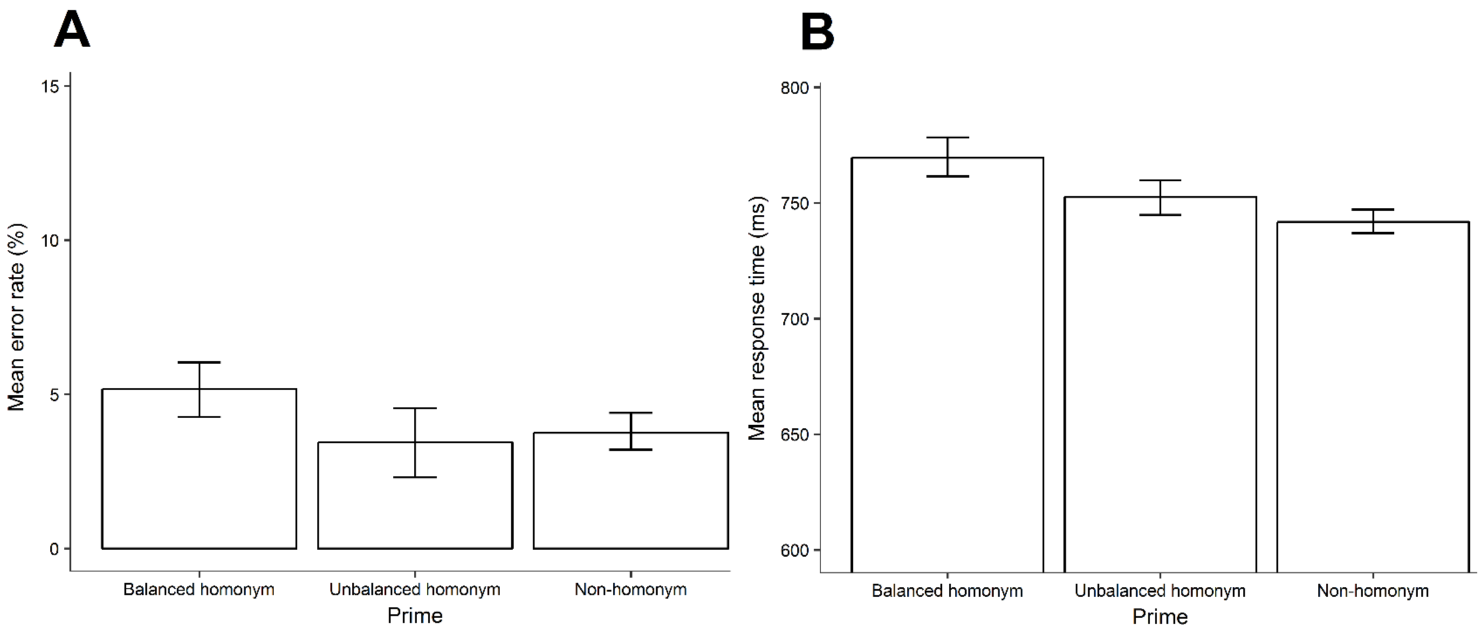

Figure 3.5: Experiment 3: Subject means of error rates (Panel A) and untransformed RTs (Panel B) across unrelated word pairs. Error bars show 95 % confidence intervals adjusted to remove between-subjects variance.

The accuracy model for related trials (marginal R^2 = .22, conditional R^2 = .48) revealed a significant Experiment × Prime interaction [$\chi^2(3) = 30.9$, $p < .001$]. The effects of balanced and unbalanced homonymy in Experiment 3 (balanced homonyms: 24.9% difference, relative to non-homonyms; unbalanced homonyms: 34.6% difference) were greater (both *p*s < .01) than those in Experiment 2 (balanced homonyms: 20.8% difference; unbalanced homonyms: 25.7% difference). There was also a marginal Experiment × Target interaction [$\chi^2(1) = 3.7$, $p = .05$], such that the latter (i.e., higher error rates for LF-meaning targets) was marginally greater ($p = .06$) in Experiment 3 (21.4% difference) than Experiment 2 (18.8% difference).

The latency model for related trials (marginal R^2 = .18, conditional R^2 = .52) revealed a significant main effect of Experiment [$\chi^2(1) = 23.4$, $p < .001$], such that responses were generally faster in Experiment 2 (*M* = 616.3 ms, 95% CIs: 586.0, 648.2) than Experiment 3 (*M* = 750.9 ms, 95% CIs: 716.3, 787.1). Further analysis showed a significant Experiment × Target interaction [$\chi^2(1) = 11.6$, $p < .001$]. The simple effect of Target (i.e., slower responses to LF- than HF-meaning targets) was greater ($p < .001$) in Experiment 3 (58.3 ms difference) than Experiment 2 (26.4 ms difference). There was also a significant Experiment × Target × Prime interaction [$\chi^2(3) = 12.7$, $p < .01$], indicating that the ambiguity effect for unbalanced homonyms in the LF meaning was greater ($p < .01$) in Experiment 3 (155.4 ms difference, relative to non-homonyms) than Experiment 2 (88.3 ms difference). All of these results remained significant in the analysis with covariates (i.e., age of acquisition, imageability, and prime-target relatedness at the target level).

3.3.2.4 Experiments 2 & 3 Combined: Unrelated Trials

Turning to unrelated trials, the response-accuracy model (marginal R^2 = .04, conditional R^2 = .32) revealed higher error rates in Experiment 2 (*M* = 8.2%, *SD* = 6.8) than Experiment 3 [*M* = 4.0%, *SD* = 3.0; $\chi^2(1) = 5.9$, $p < .05$].

The main effect of Experiment was also significant [$\chi^2(1) = 13.2$, $p < .001$] in the response-latency model (marginal R^2 = .11, conditional R^2 = .53). Responses were slower in Experiment 3 (*M* = 734.3 ms, 95% CIs: 697.1, 773.6) than Experiment 2 (*M* = 631.8 ms, 95% CIs: 597.9, 667.7). All other effects involving Experiment were not significant.

3.3.3 Discussion

Experiment 3 provided further evidence that relative meaning frequency modulates the impact of homonymy. A processing cost for balanced homonyms appeared both on related and unrelated trials and was comparable to that observed in Experiment 2. In contrast, unbalanced homonyms were again comprehended as fast as their unambiguous counterparts on both HF-meaning and unrelated trials.

Experiment 3 did not find any evidence of competition for unbalanced homonyms (i.e., ambiguity effects across all trial types) even when there was additional time to allow for LF-meaning activation. In fact, the longer prime duration in Experiment 3 seems to have weakened activation of the LF meaning even further. Error rates for unbalanced homonyms in the LF meaning were higher in Experiment 3 than Experiment 2, which points to decay or suppression of the alternative meaning over the course of 700 ms. Activation of the LF meaning was so weak toward the end of the prime presentation that most participants failed to select the meaning upon seeing a supporting target word (Frost & Bentin, 1992; MacGregor et al., 2015; Simpson & Burgess, 1985). When participants did disambiguate these words, the processing cost was greater in Experiment 3 than Experiment 2, which, again, seems to indicate more effortful retrieval and integration of the alternative meaning. Overall, then, the findings suggest that unbalanced homonymy does not produce semantic competition or response conflict due to weak activation of the LF meaning in the absence of supporting context.

Analyses contrasting the effects of prime duration also showed that the high proportion of errors on the related trials in Experiment 2 was not due to the fast-paced nature of the task or a speed-accuracy trade-off. The longer prime duration in Experiment 3 slowed participants' responding, but it did not benefit accuracy on related trials. Thus, it seems that participants had slight difficulties in detecting and judging the relatedness between the prime and the target due to the multiple constraints applied when compiling the stimuli. In particular, the targets used in these experiments were semantic (e.g., "tap-sink") rather than lexical associates[18] (e.g., "tap-water"), such that participants had to retrieve and consider a number of potentially relevant features of the word referents, which

[18] BNCweb (CQP-edition; Hoffmann & Evert, 2006) was used to explore how often the prime and the target co-occurred within spoken and written sentences, up to four words apart, in the 100-million-word BNC (2007). The analysis showed that the vast majority of the related targets were rarely used together with the primes in natural discourse, and that the two were not lexical associates. The number of sentences in which the prime and the target co-occurred ranged from 0 to 73 (M = 4.2, SD = 9.4), with only three pairs of words being relatively common ("void-null", "pen-ink", "dawn-dusk").

aimed to make the task more sensitive to the impact of semantic activation (Lucas, 2000; Thompson-Schill, Kurtz, & Gabrieli, 1998; Yee, Overton, & Thompson-Schill, 2009; see also Witzel & Forster, 2014). Furthermore, the targets were carefully selected and matched across 16 sets of word pairs based on their ambiguous status as well as 14 lexical and semantic properties (see Table 3.2 above) that have been demonstrated to influence on-line language processing. Taken together, it appears that the relatively high errors rates in Experiments 2 and 3 reflect less salient semantic relatedness of the word pairs used[19] – a by-product of rigorous control over the properties of the targets and their relation to the primes.

Although the stimulus-selection procedure seems to have increased error rates for both homonyms and non-homonyms, it was instrumental for the design of this research. Having matched the sets of related and unrelated 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), allowed for direct comparisons of ambiguity effects in different contexts/prime-target combinations. This approach, deemed essential to all tasks involving multiple stimuli, made it possible to draw reliable conclusions about the impact of prime-word manipulation that has not been contaminated by target-word variability.

---

[19] Support for this claim comes from the observation that, on average, only two of the 28 pairs in each set (range across all related sets: 0-4) 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 the primes and targets, regardless of the condition, did not elicit each other’s meanings in a typical, straightforward way.

3.4 General Discussion

The work reported in this chapter demonstrates that meaning frequency modulates the ambiguity disadvantage, such that the slowing in comprehension is restricted to balanced homonymy. This shows that any adequate explanation of ambiguity effects must take into account how meaning activation arises and changes over time due to contextual constraints and meaning frequency (see also Armstrong & Plaut, 2016). The findings for balanced and unbalanced homonymy and their implications for the different accounts of the ambiguity disadvantage are discussed in turn.

3.4.1 Unbalanced Homonymy

The findings demonstrate that unbalanced homonymy does not entail semantic competition or response-conflict resolution, most likely due to weak and delayed activation of the LF meaning out of context (Frost & Bentin, 1992; Klepousniotou et al., 2012; Meade & Coch, 2017; Simpson & Burgess, 1985). Experiments 2 and 3 showed that participants rarely retrieved the alternative meaning of unbalanced homonyms, and that a substantial processing cost arose only when they did. In other words, unbalanced homonymy slowed comprehension only when the reading of a target instantiating the unexpected LF meaning made participants engage in effortful retrieval of that meaning and suppression of the initially selected HF meaning. This interpretation is in line with Duffy et al. (1988) and Rayner and Duffy's (1986) findings. Their eye-movement studies showed that participants were slower to read LF-meaning context following an unbalanced homonym, rather than the word itself (e.g., "We knew the boxer was barking all night"), which has been taken as evidence for integration/reinterpretation processes (Leinenger, Myslín, Rayner, & Levy, 2017; Mason & Just, 2007; Musz & Thompson-Schill, 2017; Rodd, Davis, & Johnsrude, 2005; Rodd, Longe, Randall, & Tyler, 2010; Vitello et al., 2014; Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007).

Furthermore, the findings for unbalanced homonyms demonstrate that not only does meaning frequency influence meaning activation, but it also determines the outcome of word disambiguation in the absence of contextual bias. Unlike previous studies of sentence reading (Brocher et al., 2018; Duffy et al., 1988; Leinenger et al., 2017), in which strong context following an unbalanced homonym helped to ultimately retrieve the LF meaning, the current experiments show, for the first time, that readers normally fail to retrieve that meaning and correctly interpret the word in tasks with minimal context, such as relatedness decision making. Analyses of accuracy showed that, on average, participants disambiguated only 20% (Experiment 3) or 30% (Experiment 2) of the unbalanced homonyms in the LF meaning, which suggests that the meaning was activated to such a small extent that even later disambiguating information (i.e., the target word) did not lead to its selection. The view that the LF meaning of unbalanced homonyms is not fully retrieved out of context is consistent with the results of the stimulus rating studies conducted prior to Experiment 2 (see 3.2.1.2 Stimuli in this chapter). These studies showed that raters failed to detect the semantic relatedness between most of the unbalanced homonyms and their LF-meaning targets, unless their interpretation of the homonyms was primed by strong sentential context supporting the LF meaning. Overall, then, it appears that naturalistic and elaborate context is necessary to fully retrieve and select the highly uncommon meaning of an ambiguous word, both on-line and off-line.

It is important to note that the difficulty in processing the LF meanings of unbalanced homonyms did not arise because participants did not know these meanings. The norms of meaning frequency from Experiment 1 indicate that over 75% of native speakers of British English have used and/or encountered the LF meanings, which suggests that, for most participants in Experiments 2 and 3, these meanings were indeed stored in their mental lexicon. In addition, the researcher (GM) examined whether there was any relationship between the number of errors for the LF meanings of unbalanced homonyms in the present experiments and the number of participants (out of 100) in the norming experiment (Experiment 1) who reported to not know the meanings. There were no significant correlations between the two measures for both Experiment 2 (r_s = .03, p = .88) and Experiment 3 (r_s = .05, p = .82). Taken together, the implication is that unbalanced homonyms on HF-meaning and unrelated trials were comprehended as fast as non-homonyms not because participants did not know the LF meaning and processed the words as if they were unambiguous, but because the LF meaning was not sufficiently activated, and in many cases not fully retrieved, to compete with the HF meaning.

3.4.2 Balanced Homonymy

Balanced homonymy incurred a significant processing cost regardless of whether the target instantiated the slightly more frequent or the slightly less frequent meaning. The greater ambiguity effect for the latter meaning, both in terms of response accuracy and latency, should come as no surprise. Balanced homonymy is, at best, very rare in English (Armstrong et al., 2012), hence the HF and LF meanings of the balanced homonyms in the current experiments slightly differed in terms of relative frequencies (*M* = 20% difference, *SD* = 12; e.g., 60% vs. 40%). Given the chance level of performance for balanced homonyms in the LF-meaning condition, it is plausible that, for some of the words, the less frequent meaning was not sufficiently activated or retrieved to produce maximal competition. Likewise, the greater slowing in the LF-meaning condition suggests that, for these moderately balanced homonyms, selection or full retrieval of the less frequent meaning may have required suppression of the more frequent meaning, similar to that for unbalanced homonyms discussed above. Overall, then, it appears that readers are slightly biased toward the more frequent interpretation even for balanced homonyms whose meanings barely differ in relative frequency. Strong support for this claim comes from the finding that meaning dominance (β) for balanced homonyms in the current experiments was a strong predictor of both error rates and RTs for these words in the LF-meaning condition, such that the larger the difference in meaning frequencies for a balanced homonym, the greater the difficulty in responding to that word in the LF meaning. This was true for Experiment 2 (error rates: $r_s = .47$, $p < .05$; RTs: $r_s = .45$, $p < .05$) as well as Experiment 3 (error rates: $r_s = .51$, $p < .01$; RTs: $r_s = .46$, $p < .05$).

While it is possible that the greater slowing for balanced homonyms in the LF meaning may underlie additional meaning-suppression processes, this interpretation cannot accommodate the slowing on HF-meaning and unrelated trials. The effects of ambiguity on those trials were numerically similar in both experiments. Since the processing of ambiguous words on unrelated trials does not entail response conflict or reinterpretation, the findings suggest that the disadvantage effect for balanced homonyms may indeed be due to semantic competition, with additional computations involved in the processing of their less frequent meaning, such as reinterpretation or suppression of the markedly more frequent counterpart.

3.4.3 Ambiguity Disadvantage

The present findings have important implications for existing accounts of the ambiguity disadvantage. In particular, explaining why meaning frequency would modulate the effect seems to present no challenge to PDP models that predict semantic competition in comprehension. Within the PDP 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 both the speed and outcome of form-to-meaning mapping. The HF meanings of unbalanced homonyms (e.g., “money bank”), such as those used in Experiments 2 and 3, develop strong connections, and as such are activated so fast that they escape competition with the LF meanings. For balanced homonyms (e.g., football/electric fan”), meaning frequency plays barely any role in form-to-meaning mapping; both meanings are activated in parallel and to the same extent, with each being equally likely to win competition for further activation. Therefore, PDP models of ambiguity processing, such as the one implemented by Kawamoto (1993), can easily account for the impact of meaning frequency by modifying the weights on the connections between orthographic and semantic units to capture the differential effects of balanced and unbalanced homonymy on semantic activation and competition involved.

The findings are, on the other hand, inconsistent with the decision-making account whose main premise is the null effect of ambiguity on unrelated trials of the relatedness decision task (Pexman et al., 2004). Experiments 2 and 3 showed that the ambiguity effect on unrelated trials is robust but restricted to balanced homonymy, which suggests that meaning frequency may be key to explaining the inconsistencies in findings. Pexman et al. (2004) examined the effects of homonymy on unrelated trials in three relatedness decision studies that involved different sets of homonyms but, critically, neither manipulated nor controlled for the relative frequencies of their meanings. In order to determine the numbers of balanced and unbalanced homonyms in two of the three studies[20], the researcher (GM) used Twilley et al.'s (1994) norms of meaning frequency in Canadian English – the dialect spoken by Pexman et al.'s (2004) participants. As expected, Pexman et al.'s (2004) stimulus list confounded balanced and unbalanced homonymy; approximately half of their ambiguous words had unbalanced meaning frequencies that differed by 41-79%. Taken together, it seems that Pexman et al. (2004) failed to detect a reliable ambiguity effect on unrelated trials because they did not consider the role of meaning frequency.

While the decision-making account (Pexman et al., 2004) predicts semantic ambiguity to slow relatedness decisions on related but not unrelated trials because only the former involves a response conflict, the present work demonstrates that this is not the case. Experiments 2 and 3 showed consistent

[20] One of Pexman et al.'s (2004) experiments (Experiment 5) involved Katakana words, hence it was difficult to inspect the ratio of balanced and unbalanced homonyms in that stimulus list.

evidence that the processing disadvantage for balanced homonyms appeared even on unrelated trials that are free of such a conflict. Consistent with Duffy et al. (1988) and Rayner and Duffy (1986), the finding that the disadvantage effect is sensitive to meaning frequency, and its influence on semantic activation, further suggests that the effect arises due to semantic activation, rather than response-selection, processes. It appears that, as predicted by PDP models (e.g., Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004), ambiguity produces competition during semantic activation, and this competition is greater for words whose semantic representations are activated in parallel. Overall, then, Pexman et al.'s (2004) proposal that the disadvantage effect reflects response-conflict resolution that arises on related but not unrelated trials is unsupported - it cannot accommodate findings when meaning frequency is taken into account.

Although the present findings provide strong evidence against Pexman et al.'s (2004) proposal, it is possible that the effect of balanced homonymy in Experiments 2 and 3 was due to other decision-making demands. In particular, Hino et al. (2006) suggested that the processing cost for ambiguous words arises only when a task-relevant response requires analysis of their multiple meanings. It is reasonable to assume that such analysis was in play in the current experiments. In order to make a correct response on unrelated trials, participants may need to ensure that the prime is indeed unrelated to the target, and such "checking" might take longer for words with two activated meanings (i.e., balanced homonyms) than for those with one (i.e., unbalanced homonyms, non-homonyms). However, while this interpretation can accommodate finding of a processing disadvantage for balanced homonymy in Experiments 2 and 3, it fails to accommodate the same finding in Duffy et al. (1988) and Rayner and Duffy's

(1986) studies in which balanced homonymy slowed fixations on the ambiguous word in late-disambiguation sentences. That is, the idea that the disadvantage lies in the analysis of multiple meanings in relation to response selection (Hino et al., 2006) cannot explain why the effect appears in tasks that do not involve response making, unless one assumes that readers explicitly decide as to which meaning of a homonym should be selected for sentential integration, all within 200-300 ms of reading the word. Further evidence against this idea, as well as the decision-making account in general, is presented in Chapter 4.

3.4.4 Conclusions

In summary, this chapter demonstrates that homonymy slows word comprehension in minimal context, and this effect is more pronounced for homonyms with similar meaning frequencies. The most important contribution is the finding that the slowing arises both on related and unrelated trials of the relatedness decision task. This lends support to PDP models predicting a processing disadvantage ensuing from multiple form-to-meaning mappings (e.g., Rodd et al., 2004), particularly those that incorporated an explanation for the impact of meaning frequency (Kawamoto, 1993). Ambiguity produces competition for semantic activation that slows comprehension, as long as the different meanings of the ambiguous word are of comparable frequency, and are thus activated in parallel. In contrast, the findings challenge the alternative account which suggests that the disadvantage is due to task-specific decision making, such as response-conflict resolution on related, but not on unrelated, trials (Pexman et al., 2004). This account faces a daunting task of explaining the inconsistencies in findings and needs to recognise the crucial role of meaning frequency in ambiguity processing.

Indeed, this chapter shows that the distinction between balanced and unbalanced homonyms is an important one. The latter are seemingly processed as the unambiguous counterparts, and as such show weak to no effects of ambiguity. This supports a relatively recent assumption in the literature that meaning frequency modulates ambiguity effects in word processing (Armstrong et al., 2012; Brocher et al., 2018; Grindrod et al., 2014; Klepousniotou & Baum, 2007; Mirman et al., 2010). Furthermore, both Experiments 2 and 3 showed that not only does meaning frequency affect the time-course and level of meaning activation, but it also determines the outcome of the disambiguation process in minimal context. While the former was suggested as early as the ‘70s (e.g., Hogaboam & Perfetti, 1975), the current experiments reveal, for the first time, that readers quickly disambiguate unbalanced homonyms toward their highly dominant meaning, such that they often fail to correctly interpret or reinterpret the words when the alternative meaning is actually needed for successful comprehension. This is not surprising given that frequency, or dominance, has a pervasive influence on the processing of a range of linguistic materials, including words (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Inhoff & Rayner, 1986; Marslen-Wilson, 1990), syntactic ambiguities (e.g., Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001; Juliano & Tanenhaus, 1994; Trueswell, 1996), antonyms (Kostić, 2015; Ingram, Hand, & Maciejewski, 2016), as well as non-literal expressions (e.g., Filik & Moxey, 2010; Giora, 1997; Mashal, Faust, Hendler, & Jung-Beeman, 2008).

Chapter 4: Semantic Competition in Word Comprehension: EEG Evidence from Homonymy

4.1 Introduction

The work discussed in Chapter 3 suggests that the disadvantage effect is due to semantic competition, albeit that the degree of this competition is modulated by meaning frequency. Experiments 2 and 3 revealed a processing cost for balanced homonyms on both related and unrelated trials, which challenges the decision-making account (Pexman et al., 2004) and its premise that ambiguity effects appear exclusively on related trials due to response conflict. In contrast, unbalanced homonyms incurred a processing cost only when participants disambiguated the words towards the LF meaning, indicating more effortful retrieval and integration of that meaning (e.g., Duffy et al., 1988; Musz & Thompson-Schill, 2017; Sereno et al., 2006; Vitello et al., 2014), rather than semantic competition.

Although these findings are compatible with the semantic competition account (e.g., Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004), the inferences about the effects of ambiguity in Experiments 2 and 3 were largely based on differential patterns of relatedness decisions. Thus, various other processes (e.g., conflict at the response level or meaning suppression) may have also contributed to these effects. In order to validate the semantic competition account, Experiment 4 involved the same task and design as Experiment 3, but its primary aim was to determine electrophysiological indices of the effects reported in Experiments 2 and 3. Unlike RTs that are a cumulative result of a host of perceptual, cognitive, and motor processes, ERPs provide a measure of real-time language processing (see Kutas & Van Petten, 1994; Sereno & Rayner, 2003) that can establish when the ambiguity effect arises.

Therefore, Experiment 4 was designed to test the hypothesis that the processing disadvantage for balanced homonymy arises due to competition during semantic activation. Under this account, one would expect to observe the effect in the N400 component that has been linked to the ease of semantic processing (for a detailed review, see Federmeier & Laszlo, 2009). In short, the N400 refers to a negative-going wave that typically peaks 400 ms after the presentation of words, pictures, and other meaningful stimuli. Semantic priming (e.g., Brown & Hagoort, 1993; Rossell, Price, & Nobre, 2003), prior context (e.g., Nieuwland & van Berkum, 2006; van Berkum, Hagoort, & Brown, 1999), and predictability (e.g., Davenport & Coulson, 2011; Kutas & Hillyard, 1980) have all been shown to attenuate the relative amplitude of the N400 to a word. The consensus is that the component indexes semantic activation, with larger amplitudes indicating more effortful form-to-meaning mappings (for a review, see Kutas & Federmeier, 2011). Therefore, if balanced homonymy produces competition that hinders semantic activation, as seems to be the case based on the results in Chapter 3, Experiment 4 should show larger N400 amplitudes for balanced homonyms than non-homonyms and unbalanced homonyms. It is critical that this effect arises during the reading of the ambiguous prime word, separating early semantic activation processes during prime presentation from late reinterpretation or response-selection processes during target presentation.

The present experiment also aimed to uncover the mechanisms involved in the processing of unbalanced homonymy. Findings from Experiments 2 and 3 suggest that this form of ambiguity does not entail semantic competition due to weak and delayed activation of the alternative meaning in minimal context. To provide further support for this account, Experiment 4 compared the amount of semantic priming for the HF and the LF meaning of balanced and unbalanced homonyms, focusing again on the N400. The semantic priming literature has shown that targets (e.g., “tiger”) preceded by related primes (e.g., “lion”) elicit smaller N400 amplitudes than those preceded by unrelated primes (e.g., “jug”). This so-called “N400 priming effect” is thought to reflect pre-activation of the target’s meaning during the processing of the semantically related prime (e.g., Bentin, McCarthy, & Wood, 1993; Brown & Hagoort, 1993; Kutas & Hillyard, 1989). Thus, ERP studies of semantic ambiguity have often used this paradigm to explore patterns of meaning activation in homonym processing, both in isolation (e.g., Atchley & Kwasny, 2003; Klepousniotou et al., 2012; MacGregor et al., 2015; Meade & Coch, 2017) and context (e.g., Dholakia, Meade, & Coch, 2016; Elston-Güttler & Friederici, 2007; Lee & Federmeier, 2011; Kotchoubey & El-Khoury, 2014; Swaab et al., 2003). For example, a recent study by Meade and Coch (2017) found that while both meanings of unbalanced homonyms were partially activated at the SOA of 250 ms, the N400 priming effect (i.e., smaller N400 amplitudes relative to unrelated targets) was significantly greater for targets related to the HF than the LF meaning, which is consistent with the proposal that meaning frequency plays an important role in meaning activation, especially when ambiguous words are encountered on their own (for a review, see Twilley & Dixon, 2000).

Drawing on this literature, the current experiment examined the N400 to related and unrelated targets to determine the extent to which the meanings of homonymous primes are activated during relatedness decision making. For balanced homonyms, there should be a comparable N400 priming effect for targets instantiating the slightly more frequent and the slightly less frequent meaning, which would indicate that both meanings are activated in parallel and to the same extent. 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 result would support the idea, put forward in Chapter 3, that, in neutral context, readers fail to comprehend unbalanced homonyms in the unexpected alternative meaning due to reduced and insufficient activation of that meaning. In other words, for ease of comprehension, the language system is likely to process homonyms with highly uncommon meanings as functionally unambiguous words.

In summary, Experiment 4 aimed to test the hypothesis that the effect observed for balanced homonyms arises during the processing of the ambiguous word itself and is associated with the N400 component that indexes semantic processing. Such a finding is critical to explaining the effect in terms of competition during semantic activation (e.g., Kawamoto, 1993; Rodd et al., 2004), rather than in terms of decision making (Pexman et al., 2004) or analysis (Hino et al., 2006) during response selection. A subsidiary aim was to confirm, using electrophysiological evidence, that balanced and unbalanced homonyms differ in the extent to which their meanings are activated, and thus entail different levels of semantic competition.

4.2 Method

4.2.1 Participants

A group of 34 students and members of staff [27 females, aged 18-33 (*M* = 20.9, *SD* = 3.5)] from the University of Leeds participated in exchange for six course credits or £8. No participant was recruited or took part in any other study reported in this thesis. 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.

4.2.2 Stimuli & Procedure

Experiment 4 involved the same stimuli as Experiments 2 and 3, but there were three minor changes to the procedure. First, participants made relatedness decisions (for the rationale behind the task, see 1.5 Aims of the Thesis in Chapter 1) with a computer mouse, rather than a keyboard. Using their preferred (right) hand, they pressed the right key for unrelated word pairs and the left key for related word pairs. Secondly, there were four, as opposed to two, one-minute breaks – one before each experimental block (where each block included the same prime but in a different prime-target combination, as in Experiments 2 & 3). Finally, the trial procedure (shown in Figure 4.1 below) was the same as in Experiment 3, except for two adaptations. A longer ITI (1000 ms vs. 100 ms) was used that allowed participants to blink and rest their eyes, and there was a 200 ms delay between the target and response execution that aimed to minimise any overlap in ERP components evoked by the trial events. All participants were tested in a single 1.5-hour session. On average, the task lasted for 34 minutes.

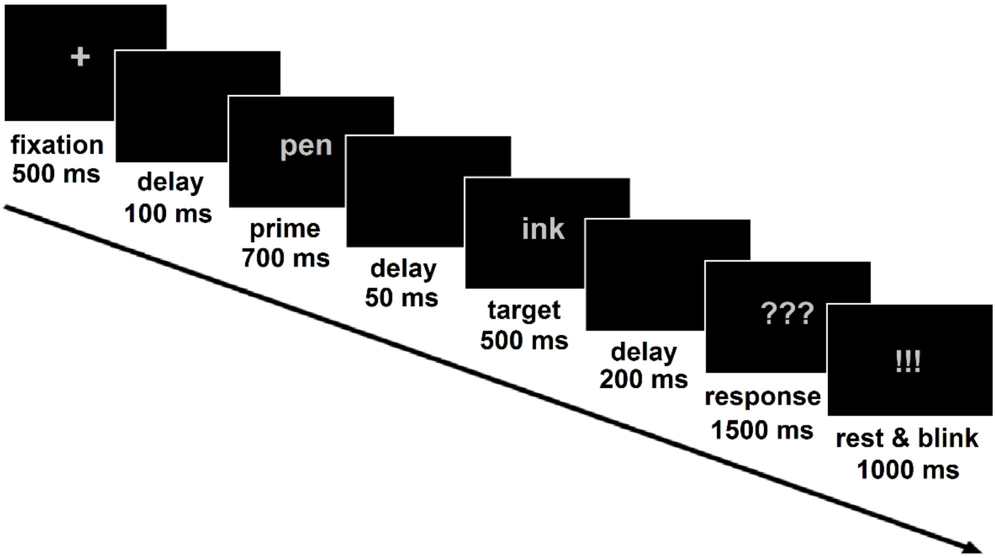

Figure 4.1: Experiment 4: Illustration of the trial sequence.

4.2.3 Data Acquisition

The electroencephalogram (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 (BioSemi, Amsterdam, the Netherlands). Recording involved 10 midline electrodes and 27 electrodes placed over each hemisphere (for the layout of the electrodes, see Figure 4.2 below). Ground electrodes were placed between Cz and CPz. Eye movements were monitored using four flat-type 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 flat-type 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.

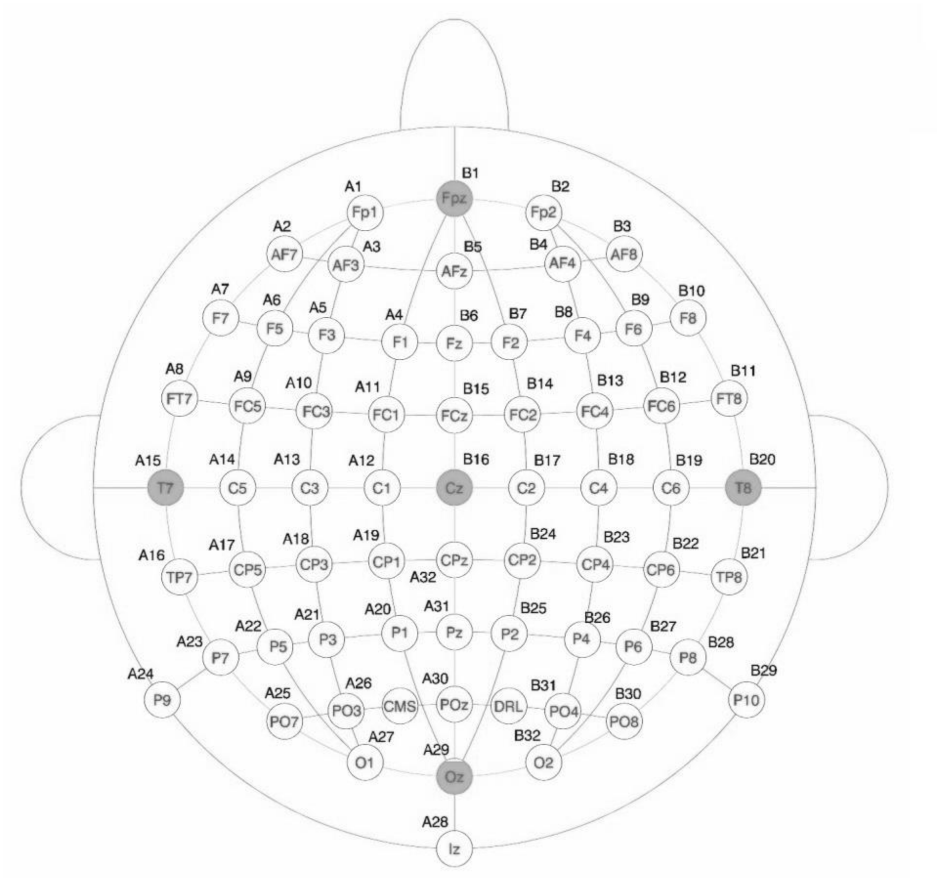

Figure 4.2: Experiment 4: Approximate layout of the 64 electrodes from which data were recorded. Image taken from http://biosemi.com/.

4.2.4 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). 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 (800 ms in total). For target-window analyses, epochs started 50 ms before the onset of the target and ended 200 ms after the offset (750 ms in total; see Figure 4.1 above). The delays of 100 ms before the onset of the prime and 50 ms before the onset of the target were used to normalise the onset voltage of the ERP waveform.

4.3 Results

4.3.1 Behavioural Data

Two of the 34 participants were removed from all analyses – one due to a large number of errors on related (37.1%) and unrelated trials (25.9%) and the other due to a large number of epochs (49.0% in the prime window, 54.6 % in the target window) containing amplifier saturation artifacts (+/- 100 µV). Accuracy data were analysed in the same way as in Experiments 2 and 3. Two separate models were conducted - one for related and one for unrelated trials. Both models included Prime and Target as fixed effects as well as random intercepts for subjects and items. Ambiguity effects in response times were this time of no interest because Experiment 4 involved a delayed response paradigm.

Mean error rates across related word pairs are illustrated in Figure 4.3 below. The results were generally similar to those of Experiments 2 and 3. The response-accuracy model for related trials (marginal R^2 = .23, conditional R^2 = .53) revealed a significant effect of Prime [$\chi^2(3) = 56.7$, $p < .001$], such that error rates were higher for both balanced (*M* = 27.2%, *SD* = 8.3; $p < .001$) and unbalanced homonyms (*M* = 43.0%, *SD* = 8.9; $p < .001$) than non-homonyms (*M* = 14.0%, *SD* = 7.0). Relatedness decisions were more erroneous to LF-meaning (*M* = 33.4%, *SD* = 8.3) than HF-meaning targets (*M* = 14.6%, *SD* = 5.6), and this effect of Target [$\chi^2(1) = 32.0$, $p < .001$] interacted with the type of Prime [$\chi^2(3) = 49.3$, $p < .001$]. Relative to both targets of non-homonyms (M_A = 11.9%, *SD* = 6.6; M_B = 14.0%, *SD* = 7.0), error rates were higher for the LF-meaning targets of balanced (*M* = 34.8%, *SD* = 12.0; $p < .001$) and unbalanced homonyms (*M* = 70.8%, *SD* = 13.9; $p < .001$), but not the HF-meaning counterparts of balanced (*M* = 19.6%, *SD* = 7.0; $p = .08$) or unbalanced homonyms (*M* = 15.2%, *SD* = 7.4; $p = .56$). None of the effects approached the significance threshold in the model for unrelated trials (marginal R^2 = .02, conditional R^2 = .35; see Figure 4.4 below).

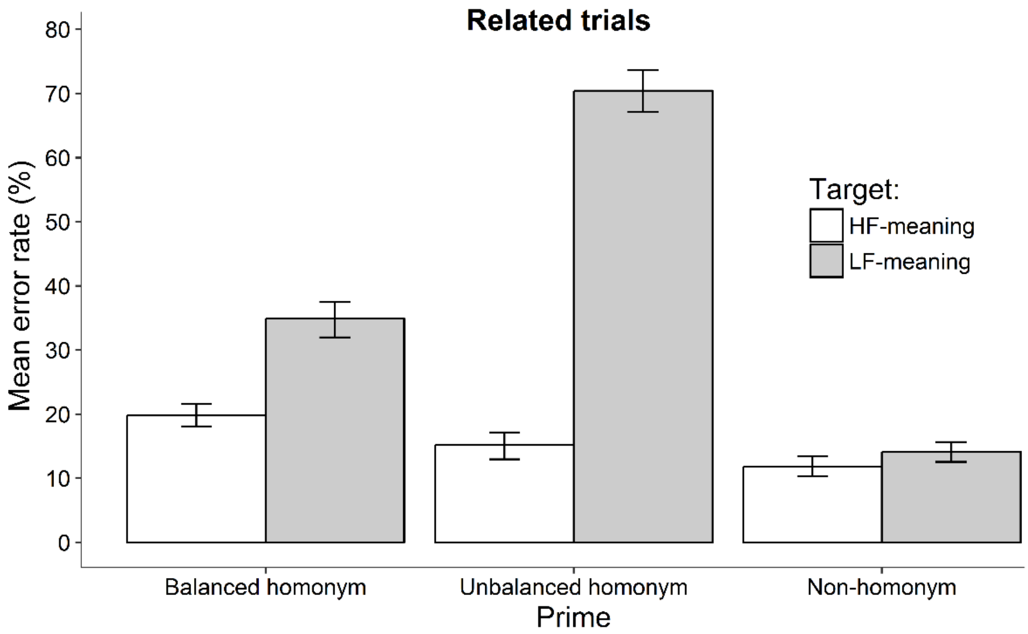

Figure 4.3: Experiment 4: Subject means of error rates across related word pairs. Error bars show 95 % confidence intervals adjusted to remove between-subjects variance.

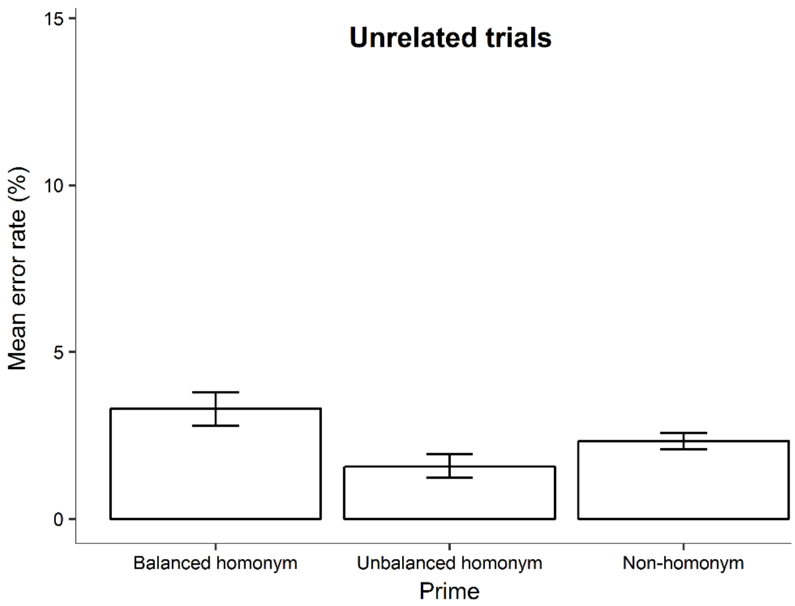

Figure 4.4: Experiment 4: Subject means of error rates across unrelated word pairs. Error bars show 95 % confidence intervals adjusted to remove between-subjects variance.

4.3.2 EEG Data

EEG analyses excluded all individual epochs that contained amplifier saturation artifacts (+/- 100 µV; 0.9% of all trials in the prime window, 1.3% in the target window) or incorrect responses (12.5% of all trials). Following recent studies (e.g., Amsel, 2011; Frömer, Maier, & Abdel Rahman, 2018; Kornrumpf, Niefind, Sommer, & Dimigen, 2016), epoched data were analysed on a trial-by-trial basis (without any prior aggregation) using linear mixed-effects modelling, primarily due to a large proportion of errors on trials involving homonyms in the LF meaning (see Figure 4.3 above). As in De Cat, Klepousniotou, and Baayen (2015), each of the 64 channels was analysed separately as there was too much data (over 700,000 observations per channel) to fit a single model. In order to prevent spurious results ensuing from a potential multiplicity problem, topographical consistency was used as an additional criterion when judging the reliability of results. The rationale was that channels are not independent so any effects specific to ambiguity should be similar across neighbouring channels.

Since the hypotheses for the prime and the target window concerned the N400, analyses focussed on the time interval that would best represent this component. Visual inspection of the waveforms during prime (see Figure 4.5 below) and target presentation (see Figures 4.7 & 4.9 below) revealed a large difference in peak latency for unbalanced homonyms during the target-word presentation (i.e., an earlier peak for the HF-meaning than LF-meaning/unrelated targets). In order to capture and account for any divergence in the waveforms, analyses of amplitudes in both the prime and the target window examined a 350-550 ms segment that was divided into four consecutive time bins of 50 ms (350-400 ms, 400-450 ms, 450-500 ms, 500-550 ms).

4.3.2.1 Prime Presentation

Prime-window analyses compared N400 amplitudes (µV) to homonyms and non-homonyms in the prime window. 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). All models included random intercepts for subjects and items as well as random by-subject slopes (mainly for the effect of Time). Planned contrasts examining the effects of Prime compared balanced/ unbalanced homonyms to both sets of non-homonyms, and their significance level was adjusted using the Holm-Bonferroni method. Only effects that involved the factor of Prime and were relevant to the hypotheses are reported below.

There was a significant interaction between the effects of Prime and Time at all channels (all *p*s < .05), except for T7, TP7, P7, and P9. However, only a subset of these channels revealed significant effects of ambiguity in pairwise comparison analyses. Amplitudes in the 400-450 ms window were larger (i.e., more negative) for balanced homonyms than non-homonyms (all *p*s < .05) 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)[21]. Relative to non-homonyms, balanced homonyms also elicited more negative amplitudes in the earlier 350-400 ms window at fronto-polar, anterio-frontal, frontal, and fronto-temporal sites (FPz, AF8, Fz, F2, F4, FT7, FT8; all *p*s < .05), as well as in the later 450-500 ms window again at similar sites (FPz, AF3, AF4,

[21] Pairwise comparisons also showed significantly larger amplitudes for balanced homonyms than non-homonyms at T8 (350-400 ms, 400-450 ms, 450-500 ms, 500-550 ms). However, this effect did not appear at neighbouring channels, and was therefore deemed unreliable.

AF8, Fz, F1, F2, F4, FT7, FT8; all *p*s < .05). There were no significant differences between unbalanced homonyms and non-homonyms (see Figures 4.5 & 4.6 below), and no other effects involving Prime. In summary, 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 both anteriorly to anterio-frontal sites and posteriorly to fronto-temporal sites.

4.3.2.2 Target Presentation

Target-window analyses compared N400 amplitudes to the related and unrelated targets of homonymous words. This involved a set of mixed-effects models with the factors of Prime (balanced homonym, unbalanced homonym), Target (HF-meaning, LF-meaning, $unrelated_{1,}$ $unrelated_{2}$), 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 investigate the amount of priming for the different meanings of balanced versus unbalanced homonyms. All models included random intercepts for subjects and items and random by-subjects slopes (most often for Block or Target). Planned contrasts examining the effects of Target 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 effects that involved Target and were relevant to the hypotheses are reported below.

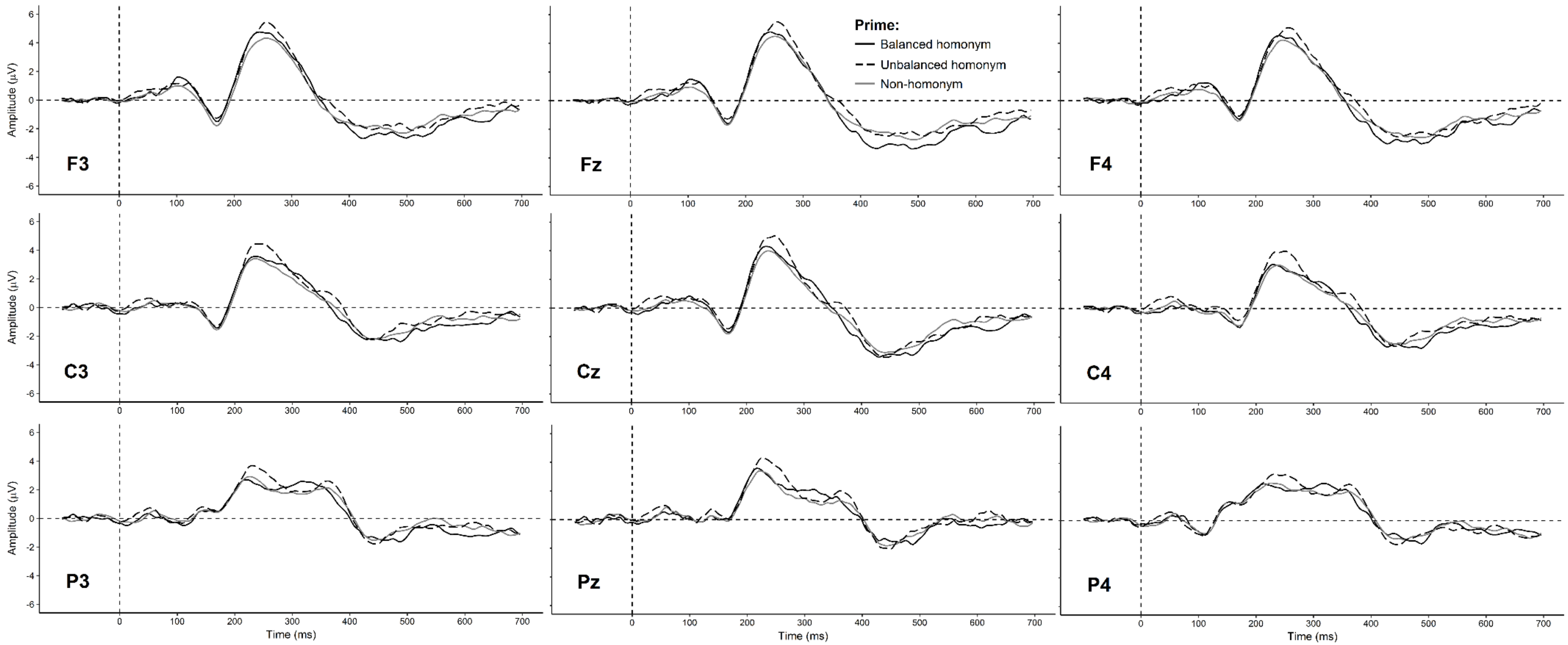

Figure 4.5: Experiment 4: Grand average waveforms for homonyms (balanced & unbalanced) and non-homonyms during prime presentation (at major frontal, central, & posterior locations). Negative amplitudes are plotted downwards.

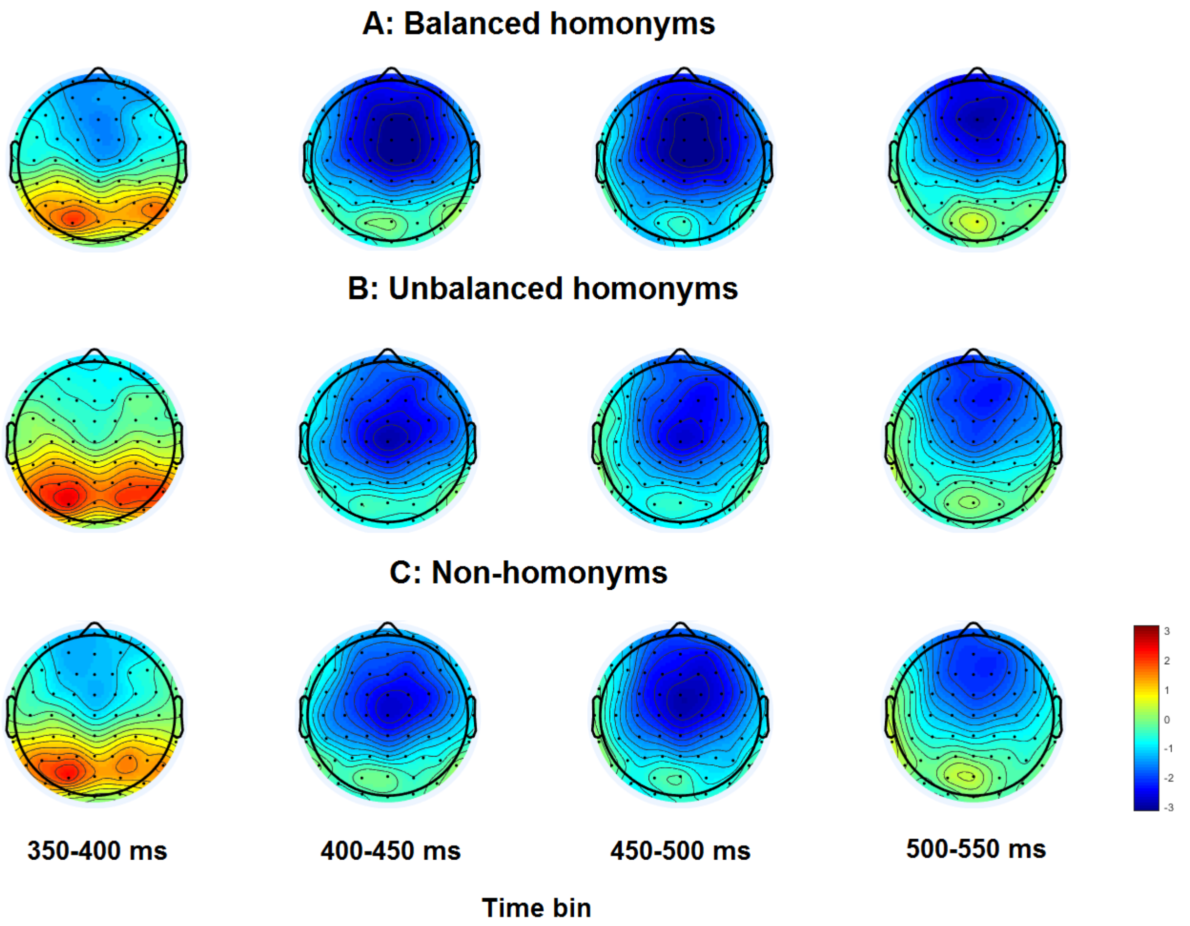

Figure 4.6: Experiment 4: Spline-interpolated topographic maps of grand average waveforms for balanced homonyms (Panel A), unbalanced homonyms (Panel B), and non-homonyms (Panel C). Isopotential line spacing is 1 µV. Same scale for all prime types.

There was a significant main effect of Target (all *p*s < .05) at several 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 for the channels showed a significant reduction in amplitudes to HF-meaning targets (all *p*s < .05) relative to unrelated targets at all the channels, as well as relative to LF-meaning targets at all the channels, except for C4, C5, CP3, CP4, CP6, P8, and PO8. There were no significant differences between LF-meaning and unrelated targets[22].

Analyses also revealed a significant Target × Time interaction (all *p*s < .05) at all channels, except for P9. The reduction in amplitudes to HF-meaning targets relative to unrelated targets was significant (all *p*s < .05) in both the 450-500 ms and the 500-550 ms window at all the channels, except for AF7, AF8, F5, F7, F8, FT7, T7, TP7, P7, and P10. This effect was also significant (all *p*s < .05) in the earlier 400-450 ms window, mainly at fronto-central, central, centro-parietal, parietal, and parieto-occipital sites (Fz, F2, FCz, FC1, FC2, FC4, T8, Cz, C1, C3, C5, C2, C4, CPz, CP1, CP3, CP5, CP2, CP4, CP6, Pz, P1, P3, P5, P2, P4, P6, POz, PO3, PO4, PO8, Oz, & O2).

The reduction in amplitudes to HF-meaning targets relative to the LF-meanings counterparts was significant (all *p*s < .05) in the 500-550 window at similar sites (Fz, F2, FCz, FC1, FC2, FC4, Cz, C1, C3, C5, C2, C4, CPz, CP1, CP3, CP5, CP2, CP4, CP6, Pz, P1, P3, P5, P2, P4, P6, POz, PO3, PO4, Oz, O1, O2, & Iz). This effect was also significant (all *p*s < .05) in the earlier windows of

---

[22] Pairwise comparisons also showed significantly smaller amplitudes to LF-meaning than unrelated targets at FT8 and CP6. However, these effects did not appear at respective neighbouring channels, and were therefore deemed unreliable.

400-450 ms and 450-500 ms at the same channels as in the 500-550 ms window, except for F2, FC4, C1, C5, C4, CP3, CP5, CP6, P2, and Iz.

Finally, there was a significant reduction in amplitudes to LF-meaning targets relative to unrelated targets in the 450-500 ms window, at a smaller cluster of centro-parietal, parietal, and parieto-occipital sites (CP4, CP6, P2, POz, & PO8[23]; all *p*s < .05).

Analyses also revealed a significant Target × Time × Prime interaction (all *p*s < .05) at all channels, except for T7, TP7, P7, and P9. For balanced homonyms (see Figures 4.7 & 4.8 below), the reduction in amplitudes to HF-meaning targets relative to the unrelated counterparts was significant (all *p*s < .05) in the last 500-550 ms window, predominantly at fronto-central, central, centro-parietal, parietal, and parieto-occipital sites (Fz, F4, FCz, FC1, FC3, FC2, FC4, Cz, C1, C3, C5, C2, C4, CPz, CP1, CP3, CP5, CP2, CP4, CP6, Pz, P1, P3, P5, P2, P4, P6, P8, POz, PO3, PO4, & O2). This effect was also significant (all *p*s < .05) in the earlier 450-500 ms window, at a smaller cluster of centro-parietal, parietal, and occipito-parietal sites (Cz, CPz, CP1, CP2, CP6, Pz, P1, P2, POz, PO3, & PO4). The contrasts between HF-meaning and LF-meaning targets as well as between LF-meaning and unrelated targets for balanced homonyms were non-significant.

For unbalanced homonyms (see Figures 4.9 & 4.10 below), on the other hand, the reduction in amplitudes to HF-meaning targets relative to the unrelated counterparts was significant (all *p*s < .05) in both the 450-500 ms and the 500-550 ms window at all the channels, except for AF7, AF8, F5, F7, FT7, and P10 (excluding T7, TP7, P7, & P9 that did not reveal the 3-way interaction in the first

---

[23] Pairwise comparisons also showed smaller amplitudes to LF-meaning than unrelated targets at FT8 in all four time bins (350-400 ms, 400-450 ms, 450-500 ms, 500-550 ms). However, this effect was not as sustained at any other channel, and was therefore deemed unreliable.

place). This effect was also significant (all *p*s < .05) in the earlier 400-450 ms window at the same channels as in the other two windows, except for AF3, FP1, F3, FC5, TP8, CP5, P8, PO7, and Iz.

The reduction in amplitudes to HF-meaning targets relative to LF-meaning targets of unbalanced homonyms was significant (all *p*s < .05) 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). This effect was also significant (all *p*s < .05) at similar sites in the earlier windows of 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 (FPz, FP1, FP2, AFz, AF3, AF4, Fz, F1, F4, F6, FCz, FC1, FC3, FC2, FC4, Cz, C1, C2, C4, C6, CPz, CP1, CP2, Pz, P1, P3, PO3, Oz, & O2). The contrasts between LF meanings and unrelated targets were not significant.

Overall, then, 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 homonym types, priming for the HF meaning appeared over bilateral medial and lateral centro-parietal sites, extending anteriorly to frontal sites and posteriorly to occipital sites.

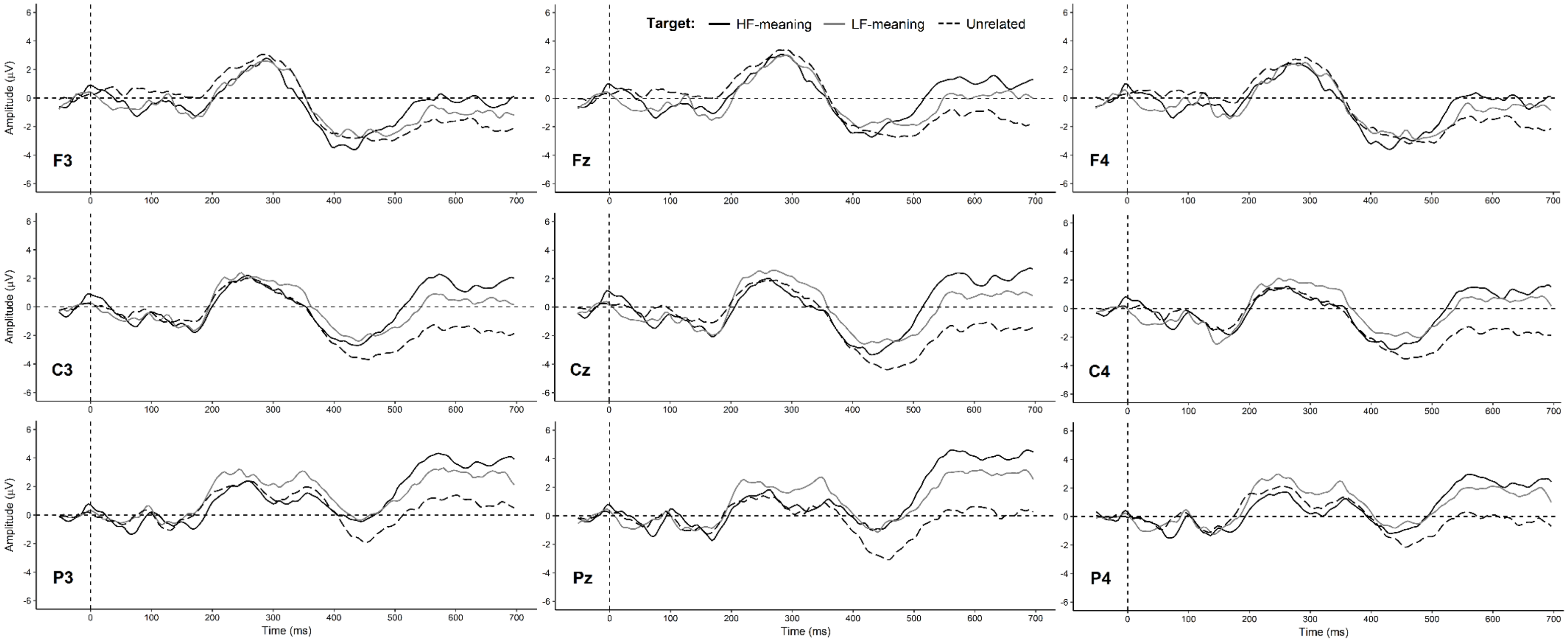

Figure 4.7: Experiment 4: Grand average waveforms for the related (HF-meaning & LF-meaning) and unrelated targets of balanced homonyms during target presentation (at major frontal, central, & posterior locations). Negative amplitudes are plotted downwards.

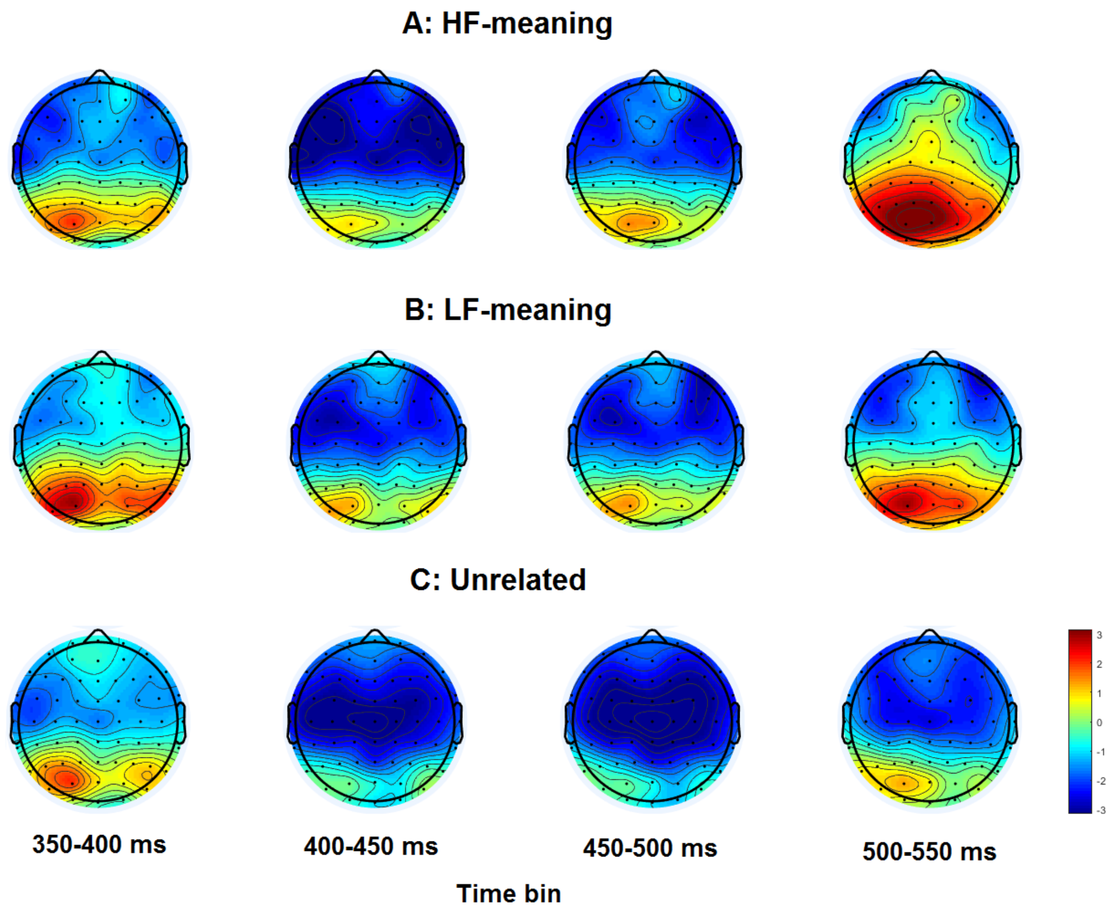

Figure 4.8: Experiment 4: Spline-interpolated topographic maps of grand average waveforms for the HF-meaning (Panel A), LF-meaning (Panel B), and unrelated targets (Panel C) of balanced homonyms. Isopotential line spacing is 1 µV. Same scale for all target types.

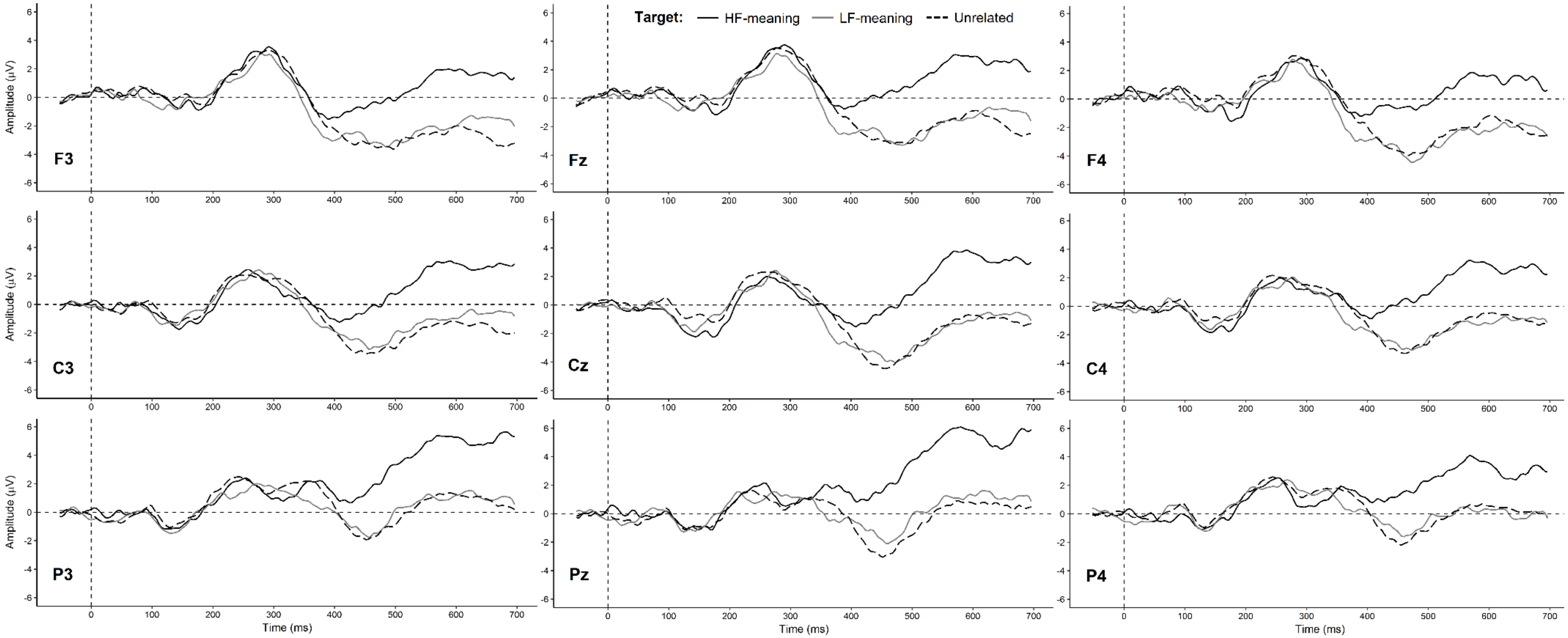

Figure 4.9: Experiment 4: Grand average waveforms for the related (HF-meaning & LF-meaning) and unrelated targets of unbalanced homonyms during target presentation (at major frontal, central, & posterior locations). Negative amplitudes are plotted downwards.

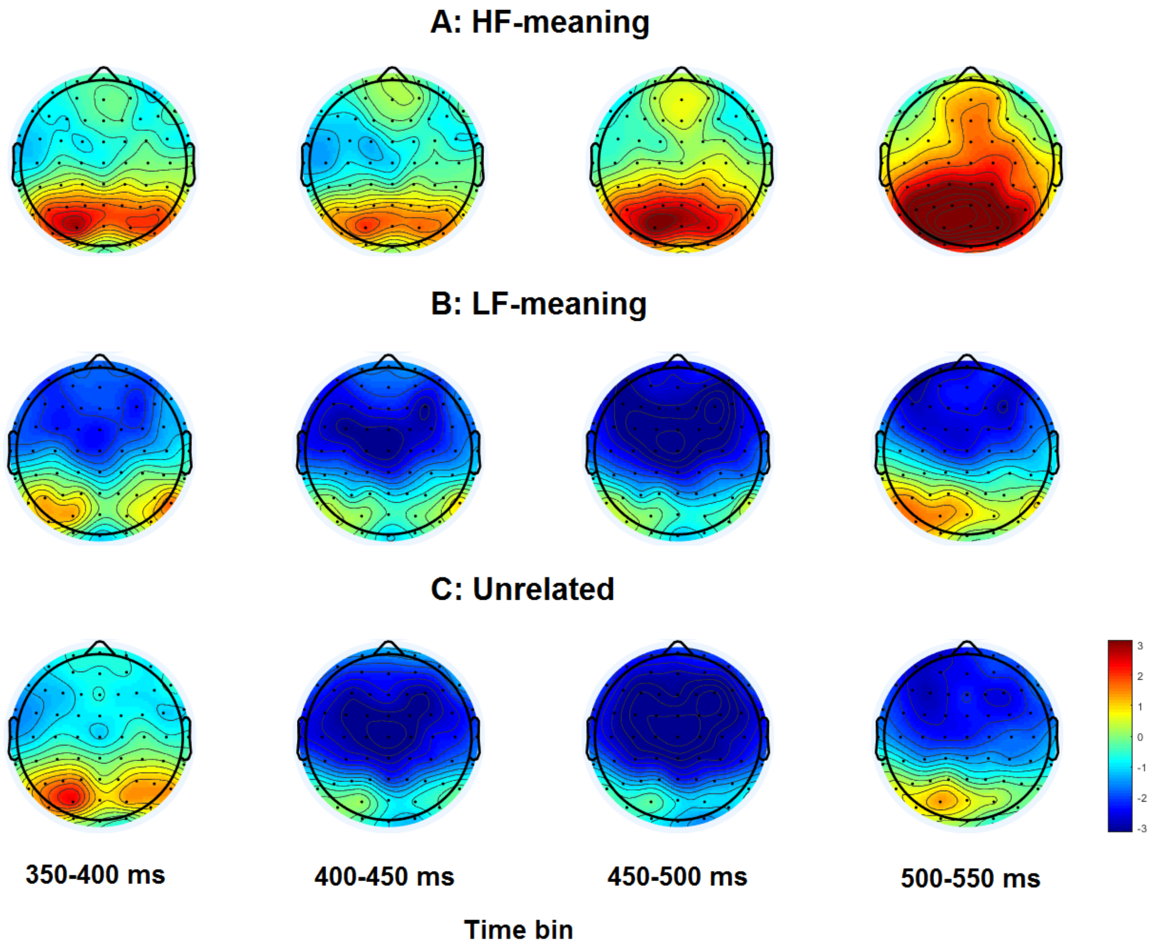

Figure 4.10: Experiment 4: Spline-interpolated topographic maps of grand average waveforms for the HF-meaning (Panel A), LF-meaning (Panel B), and unrelated targets (Panel C) of unbalanced homonyms. Isopotential line spacing is 1 µV. Same scale for all target types.

4.4 Discussion

Two key findings emerged from the current experiment. First, analyses for the prime window showed that balanced homonyms elicited significantly larger/more negative N400 amplitudes over frontal sites than non-homonyms. Second, analyses for the target window revealed substantial priming for the HF meaning of unbalanced homonyms and none for the LF meaning, in contrast to comparable priming for both meanings of balanced homonyms. These two findings and their implications for the study of the ambiguity disadvantage are discussed in turn.

4.4.1 Findings from the Prime-window Analyses: Semantic Competition

Experiment 4 showed increased frontal negativity, in the prime-window, from 350 ms to 500 ms post-prime-onset for balanced but not unbalanced homonyms. The finding that homonymy in general had an impact on the N400, as opposed to any other ERP component, is consistent with previous studies on brain responses to homonyms in lexical decision tasks (Beretta et al., 2005; Haro et al., 2017). For example, a recent ERP study by Haro et al. (2017) found that homonyms elicited larger N400 responses than non-homonyms, indicating that ambiguity affects semantic activation that the N400 is thought to reflect (for a detailed review, Kutas & Federmeier, 2011). A similar effect of homonymy was observed on the M350, a magnetoencephalographic (MEG) component considered to be equivalent to the EEG N400 component (e.g., Pylkkänen & Marantz, 2003). In their MEG study, Beretta et al. (2005) reported that the M350 to homonyms peaked later than that to the non-homonyms. The present experiment corroborates and extends these findings by demonstrating that homonymy affects the N400 component in a semantically engaging task that requires word disambiguation (i.e., relatedness vs. lexical decisions). However, the experiment also shows that it is balanced, but not unbalanced, homonymy that drives this effect. Thus, this is the first study to provide EEG evidence for the idea that meaning frequency modulates ambiguity effects (for behavioural evidence, see Experiments 2 & 3 in this thesis; Armstrong et al., 2012; Brocher et al., 2018; Duffy et al., 1988; Rayner & Duffy, 1986).

It is important to note that while the latency of the effect of balanced homonymy in Experiment 4 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 an N400 effect is normally largest over centro-parietal sites, rather than frontal sites as in the current experiment. This striking difference in topography suggests that the common explanation for an N400 effect in terms of differences in the extent of semantic activation during form-to-meaning mappings (Laszlo & Federmeier, 2011) may not hold true for the effect reported here. The present findings are instead consistent with Lee and Federmeier's (2006) semantic relatedness decision study that examined ERPs to prime words within minimal phrases that specified their syntactic class (i.e., "the" for nouns, "to" for verbs). Their main result was that amplitudes over frontal and fronto-central sites were larger/more negative to homonyms with noun/verb interpretations (e.g., "duck") than non-homonyms with noun/verb interpretations (e.g., "vote"; see also Lee & Federmeier, 2009; Mollo, Jefferies, Cornelissen, & Gennari, 2018). Although this effect was more sustained (250-900 ms) than the one in the present experiment (350-500 ms), it seems that enhanced frontal negativity reflects an additional, inhibitory process involved in semantic ambiguity resolution – most likely semantic competition, as suggested by the functional magnetic resonance imaging (fMRI) literature reviewed next.

Although the scalp topography of an ERP effect cannot be used to make reliable inferences about the localisation of neural sources, the finding of an ambiguity effect over frontal sites is in line with neuroimaging research on the processing of semantic ambiguity in neutral (e.g., Bedny, McGill, & Thompson-Schill, 2008; Bilenko, Grindrod, Myers, & Blumstein, 2009; Grindrod et al., 2014; Klepousniotou, Gracco, & Pike, 2014; Whitney, Jefferies, & Kircher, 2011) and in biasing context (e.g., Mason & Just, 2007; Rodd et al., 2005; Rodd et al., 2010; Vitello et al., 2014; Zempleni et al., 2007). In short, this body of research has demonstrated that the left inferior frontal gyrus (LIFG), in particular its middle (pars triangularis, Brodmann's area 45) and posterior areas (pars opercularis, Brodmann's area 44), is the most consistent brain region that shows an increased haemodynamic response to ambiguity (for a meta-analysis, see Rodd, Vitello, Woollams, & Adank, 2015). There also appears to be wide agreement in this literature (for a review, see Vitello & Rodd, 2015) 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 et al., 2009) or when the word must be reinterpreted following initial selection of the incorrect meaning (e.g., Vitello et al., 2014). The former situation (i.e., ambiguity out of context) closely corresponds to the prime-presentation phase in this experiment, hence the greater frontal negativity for balanced homonyms in the prime window may be the electrophysiological marker of the necessity to resolve competition between the meanings of these words within the LIFG.

This interpretation is also in line with the “conflict resolution” account of LIFG function (e.g., Novick, Kan, Trueswell, & Thompson-Schill, 2009; Novick, Trueswell, & Thompson-Schill, 2005; Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997), according to which the role of the posterior LIFG is to resolve competition between multiple representations at the conceptual, rather than the response-based, level. This dissociation is supported, for example, by the finding that manipulations of response conflict, such as those in the Go/No-Go task[24], produce consistent activation in the right, rather than the left, inferior frontal cortex (for a meta-analysis, see Nee, Wager, & Jonides, 2007). Further, under the conflict resolution account, the LIFG is assumed to resolve competition between any verbal information, consistent with the finding that increased activation in this region is associated with semantic (e.g., Bedny et al., 2008), syntactic (e.g., Fiebach, Vos, & Friederici, 2004), as well as phonetic competition (e.g., Blumstein, Myers, & Rissman, 2005). This view is particularly exemplified in Novick et al.’s (2009) proposal that the 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 this experiment produced the latter type of competition (at the semantic level), increased frontal negativity for the words may indeed reflect increased activation of the LIFG in an attempt to resolve that competition.

---

[24] The task requires participants to make a response in one condition but inhibit a response in another.

Taken together, Experiment 4 found that balanced, but not unbalanced, homonymy had a negative impact on the amplitude of the N400. This finding supports the proposal, presented in Chapter 3, that the ambiguity disadvantage is restricted to homonyms with equally frequent meanings, but it also provides important insight into the locus of the disadvantage. To begin with, the finding that the impact of balanced homonymy was evident in the prime window confirms that the ambiguity disadvantage arises during the actual processing of the ambiguous word, rather than during later stages of task performance. More specifically, it arises during semantic activation, as suggested by increased negativity in the N400 window. Critically, the impact of balanced homonymy was observed over frontal brain regions, and may therefore reflect increased LIFG activation produced by the necessity to resolve competition between equally dominant and plausible meanings.

This set of findings lends strong support to the semantic competition account of the ambiguity disadvantage (e.g., Borowsky & Masson, 1996; Rodd et al., 2004). The experiment shows that the disadvantage is due to competition that occurs during semantic activation, or form-to-meaning mapping. It also confirms that such competition occurs only for homonyms with more balanced meaning frequencies; this is further supported by the findings from the target-window analyses discussed below. Finally, the present experiment provides unequivocal evidence against the decision-making account, according to which the ambiguity disadvantage (in semantic relatedness decisions) occurs when a response to the semantically related target is consistent with one meaning of the ambiguous prime but inconsistent with the other (Pexman et al., 2004), or when a response to the unrelated target requires participants to check that all of the prime’s meanings are truly unrelated (Hino et al., 2006). The finding that balanced homonymy affected brain activity in the prime window indicates that its effect occurs during the processing of the ambiguous word itself, hundreds of milliseconds before they see the related/unrelated target that follows. In other words, it demonstrates that the ambiguity disadvantage is not due to response making difficulties upon the presentation of the target, but due to semantic competition arising during the processing of ambiguity and the decision-making account (Pexman et al., 2004; Hino et al., 2006) is unable to explain such findings.

4.4.2 Findings from the Target-window Analyses: Bias in Activation

The experiment also indicates that balanced and unbalanced homonyms differ in the extent to which their multiple meanings are activated in the absence of context. Analyses for the target window showed a significant N400 priming effect (i.e., reduced amplitudes relative to unrelated targets) for HF-meaning targets and a non-significant effect for the LF-meaning counterparts, both for balanced and unbalanced homonyms. Note, however, that there was evidence to suggest that (weak) priming also occurred for the LF meaning of balanced homonyms. The LF-meaning targets of balanced homonyms elicited amplitudes that were (a) numerically, though not statistically, smaller than those to the unrelated targets and (b) comparable to those of the HF-meaning targets (see Figure 4.7 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 meaning was activated (to a lesser degree) only for balanced homonyms.

This pattern of priming is in line with the broader literature on ambiguity resolution (for a review, Twilley & Dixon, 2000). That is, the general finding of stronger priming for the dominant meaning supports the reordered access model (Binder & Rayner, 1998; Duffy et al., 1988; Duffy et al., 2001), according to which meaning activation is exhaustive but biased toward the more frequent meaning. The finding of minimal priming for the alternative meaning should also come as no surprise. As discussed previously, research into the time-course of meaning activation out of context has shown that activation of the LF meaning is low to begin with and starts to decrease approximately 300 ms after the onset of the homonym (Frost & Bentin, 1992; Simpson & Burgess, 1985). Similar evidence of meaning decay (or suppression) comes from two ERP studies that found N400 priming for the HF and LF meanings of balanced and unbalanced homonyms at the SOA of 250 ms (Klepousniotou et al., 2012) but no priming for either meaning of the same words at the SOA of 950 ms (MacGregor et al., 2015). It appears that in the current experiment, with the SOA of 750 ms, targets related to the LF meaning received little (balanced homonymy) or no priming (unbalanced homonymy) because activation of that meaning had weakened or decayed over the long presentation of the ambiguous word. This explanation is supported by the finding in Chapter 3 that homonyms in the LF meaning were significantly more difficult to process after the long (750-ms SOA in Experiment 3) than the short prime duration (250-ms SOA in Experiment 2).

The pattern of priming in this experiment supports the interpretation of the behavioural evidence in Chapter 3. In particular, it corroborates the view that readers often fail to comprehend and correctly respond to the LF meanings of unbalanced, and to a lesser extent balanced, homonyms because these meanings are not sufficiently activated and rapidly decay in the absence of context. The LF meanings of unbalanced homonyms seem to be minimally activated even on correct trials, suggesting that the substantial slowing for these words in the LF-meaning condition in Experiments 2 and 3 may be due to the increased difficulty in retrieving the alternative meaning and suppressing the dominant meaning. In other words, the difficulty in retrieving and selecting the less frequent meaning, especially in minimal context, results from strong bias towards the more frequent meaning. The language system seems to rely quite heavily on meaning frequency, or our long-term experience with homonyms, to make fast, but often hasty and incorrect, decisions as to which meaning these words may instantiate.

In addition, the present experiment shows that such bias in meaning activation is stronger for unbalanced than balanced homonyms, as revealed by no priming for the LF meaning for the former but weak priming for the latter, even after a long delay. Note, however, that this finding contrasts with that of two previous studies in which balanced and unbalanced homonyms produced either equal priming following a short delay (250-ms SOA, Klepousniotou et al., 2012) or no priming following a very long delay (950-ms SOA, MacGregor et al., 2015). There are several differences between these studies and the current one, making it difficult to ascertain the precise cause of the divergent patterns of priming. One particularly likely explanation is that the differences in meaning activation between the two types of homonymous words are pronounced when responses to the target are made based on full retrieval of a particular meaning of the ambiguous prime (relatedness decisions in Experiment 4), but not when they simply benefit from, rather than require, such retrieval (lexical decisions in Klepousniotou et al., 2012; MacGregor et al., 2015). Although this explanation remains a conjecture, it is in line with the finding that the differential impact of balanced and unbalanced homonymy in word processing, assumed to result from those differences in activation, is observed in tasks examining word comprehension (Experiments 2 & 3 in this thesis; Duffy et al., 1988; Rayner & Duffy, 1986), but not in tasks examining word recognition (Grindrod et al., 2004; Klepousniotou & Baum, 2007; but cf. Armstrong et al., 2012). The present experiment is, therefore, the first to provide electrophysiological evidence for the long-held assumption that balanced and unbalanced homonyms differ in how their meanings are accessed (Duffy et al., 1988; Rayner & Duffy, 1986).

4.4.3 Conclusions

In summary, the work presented in this chapter provides comprehensive evidence that the ambiguity disadvantage in word comprehension is due to semantic competition (e.g., Armstrong & Plaut, 2008; Borowsky & Masson, 1996; Rodd et al., 2004). Consistent with behavioural research (Experiments 2 & 3 in this thesis; Duffy et al., 1988; Rayner & Duffy, 1986), the experiment confirms that the ambiguity disadvantage is restricted to balanced homonymy and demonstrates, for the first time, that this effect arises during the semantic processing of the ambiguous word itself. More specifically, the findings suggest that balanced homonymy produces competition during the semantic activation process which most likely engages the LIFG that has been implicated in the resolution of such competition (e.g., Bedny et al., 2008; Bilenko et al., 2009; Grindrod et al., 2014; Novick et al., 2009). Another important contribution of the current experiment is the direct demonstration that balanced and unbalanced homonyms differ in how their meanings are accessed out of context, and that this determines the degree of competition they produce. As proposed in Chapter 2, the equally frequent meanings of balanced homonyms are activated to a relatively similar extent, and therefore have the best potential to compete in the race for further activation. The less frequent meanings of unbalanced homonyms, on the other hand, do not reach sufficient activation to be able to compete with the highly frequent meanings.

The findings lend strong support to the postulate of semantic competition in PDP models of ambiguity processing (e.g., Armstrong & Plaut, 2008; Rodd et al., 2004), especially those that incorporated an explanation for the impact of meaning frequency (Kawamoto, 1993). It has become clear that the rejection of this postulate by Hino et al. (2006) and Pexman et al. (2004) is unsubstantiated, and that, in fact, it is their decision-making account that struggles to explain the impact of ambiguity, in particular as delineated in Experiments 2-4. To begin with, the account does not specify why the ambiguity disadvantage, which is assumed to be unrelated to semantic activation, would be sensitive to meaning frequency and its well-documented impact on semantic activation. Furthermore, the account fails to accommodate the current findings of competition effects for balanced homonyms arising during the semantic processing of the ambiguous word itself, rather than during the processing of related or unrelated targets and subsequent response making. Likewise, if the disadvantage effect is purely a task artefact at the response-selection stage, as Hino et al. (2006) and Pexman et al. (2004)

suggest, it remains unclear why it would repeatedly appear across a number of tasks of distinct response-selection demands. After all, competitive processes involved in understanding ambiguous words have been observed in tasks involving semantic relatedness (e.g., Gottlob et al., 1999; Hoffman & Woollams, 2015; Jager, Green, & Cleland, 2016) and categorisation decisions (e.g., Hino et al., 2002; Jager & Cleland, 2015), semantically primed (e.g., Balota & Paul, 1996; Chwilla & Kolk, 2003; Kandhadai & Federmeier, 2007) and unprimed lexical decisions (e.g., Armstrong & Plaut, 2016; Beretta et al., 2005; Tamminen et al., 2006), sensicality judgements[25] (e.g., Brown, 2008; Klein & Murphy, 2001; Klepousniotou et al., 2008), and even sentence-reading tasks that do not require any response or decision (e.g., Brocher, Foraker, & Koenig, 2016; Duffy et al., 1988; Frazier & Rayner, 1990; Rayner & Duffy, 1986). Using electrophysiological evidence, the present work marks a significant step toward unravelling the locus of competition effects in ambiguity processing, in that it clearly refutes any proposal that they are due to decision making at the response-selection stage; instead, it establishes that competition effects arise during semantic processing.

---

[25] In this task, participants decide whether two consecutive modifier-noun phrases “make sense”. The phrases instantiate the same or different interpretation of the noun (e.g., “blind date” followed by “dinner date” vs. “expiration date”).

Chapter 5: Semantic Competition in Word Comprehension: Evidence from Learning New Meanings for Known Words

5.1 Introduction

PDP models of ambiguity processing (e.g., Armstrong & Plaut, 2008; Borowsky & Masson, 1996; Kawamoto, 1993; Rodd et al., 2004) assume that ambiguous words have separate semantic representations that compete with one another during the activation process. The present work has hitherto shown that this is the case for homonyms, albeit as long as the representations map onto meanings of comparable frequency (and are thus activated in parallel). The next natural step is to establish whether the ambiguity disadvantage also arises for polysemes that have multiple related senses, rather than unrelated meanings. As shown in Chapter 1, previous attempts to do so have been limited and produced mixed results, with investigations reporting either significant (Hoffman & Woollams, 2015; Pexman et al., 2004) or null effects (Frazier & Rayner, 1990; Hino et al., 2006); hence it remains unclear whether PDP models are correct to predict competition for all ambiguous words.

Extensions of the work on the ambiguity disadvantage to polysemy are also of great value to the literature on the representation and processing of ambiguity. As evident below, there is strong disagreement in the literature on whether and how polysemy differs from homonymy, partly because previous studies adopted drastically different views of what makes a word a polyseme.

Studies of so-called “lexicographic polysemy” defined polysemy in terms of the number of word senses in a dictionary (e.g., Beretta et al., 2005; Rodd et al., 2002), whereas studies of “subjective polysemy” defined this form of ambiguity based on participants’ ratings of meaning relatedness (e.g., Klein & Murphy, 2002; Haro et al., 2017). There have also been several studies of “linguistic polysemy” whereby distinctions between homonymy and polysemy as well as between different forms of polysemy were made based on linguistic principles (e.g., Frisson & Pickering, 1999; Klepousniotou & Baum, 2007). Findings from these three lines of research, and whether they predict polysemy to entail semantic competition, are discussed in turn.

5.1.1 Lexicographic Polysemy

Research into lexicographic polysemy focussed on ambiguity effects in word recognition. Following Rodd et al. (2002), the distinction between homonymy and polysemy was made based on the Wordsmyth Dictionary (Parks et al., 1998) that has separate entries for unrelated word meanings but a single entry for semantically and etymologically related word senses. The general finding of this line of research is that polysemy facilitates lexical decision performance, such that many-sense polysemes (e.g., “belt”) are processed significantly faster than their few-sense counterparts (e.g., “ant”; Armstrong & Plaut, 2008, 2011, 2016; Beretta et al., 2005; Jager & Cleland, 2016; Rodd et al., 2002; Tamminen et al., 2006). Homonymy, on the contrary, appears to have no reliable impact. There is some indication of inhibition, but the effect is weak and observed only when using pseudo-homophonic non-words or when degrading stimulus quality, both of which are thought to make lexical decisions more difficult and reliant on semantic access (Armstrong & Plaut, 2008, 2016; Azuma & Van Orden, 1997).

To accommodate these findings, Rodd et al. (2004) implemented a PDP model that explains the relatedness effect in terms of differences in how unrelated meanings and related senses are represented and activated (for a more detailed explanation, see Rodd, 2018). At the representational level, the model assumes that both homonyms and polysemes have multiple semantic representations. However, since the different senses of polysemes share semantic features, their representations may correspond to neighbouring, rather than distant, regions in semantic space (illustrated in Figure 5.1 below), which is consistent with the long-held assumption that conceptual overlap is linked to representational proximity (e.g., Collins & Loftus, 1975). At the processing level, the model assumes that activation of separate representations is inconsistent and creates semantic competition for homonyms, both at the early (i.e., word recognition) and the late stage of processing (i.e., word comprehension). For polysemes, on the other hand, activation of overlapping representations is consistent and benefits word recognition at first, but then becomes inconsistent as readers begin to settle on a specific sense. Overall, then, this model predicts competition in word comprehension both for homonyms and polysemes, but the latter should be smaller and involve only those features that are unique to the different word referents.

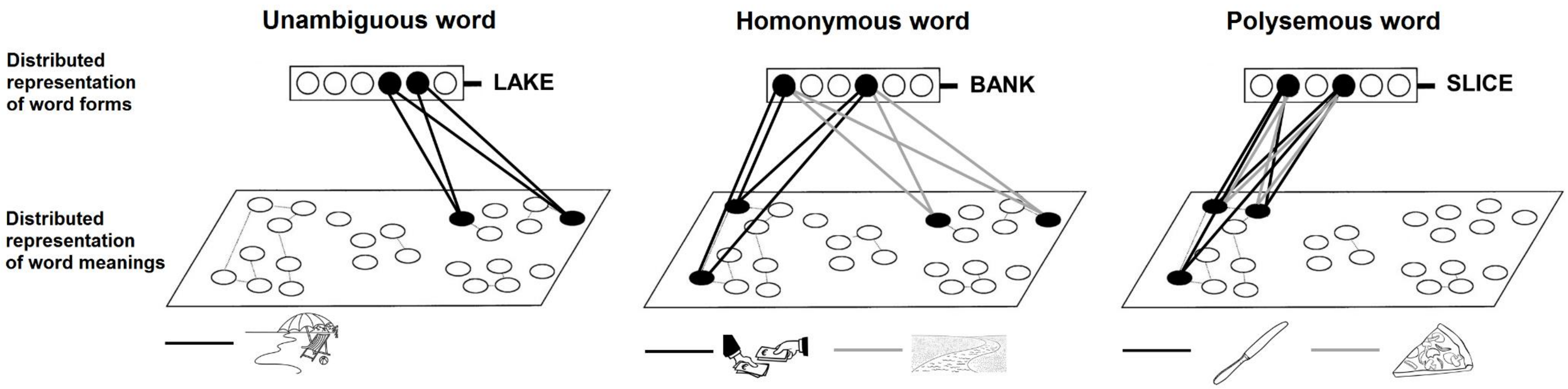

Figure 5.1: Illustration of Rodd et al.'s (2004) view of how ambiguous words are represented in the mental lexicon. Black circles in the top box represent orthographic features/units that contribute to the representation of the word form. Black circles in the bottom box represent semantic features/units that contribute to the representation of the meaning. Black and grey lines represent different form-to-meaning mappings. Images depicting different interpretations come from http://images.google.com/ (all allowed for non-commercial reuse).

5.1.2 Subjective Polysemy

The view that homonymy and polysemy differ in semantic representation and their impact in word recognition (e.g., Armstrong & Plaut, 2008; Rodd et al., 2004) is still controversial. Perhaps the leading criticism of the studies of lexicographic polysemy concerns their operationalisation of key variables. On the one hand, using dictionaries to distinguish unrelated meanings from related senses succeeded in dealing with the natural correlation between the number of interpretations and their relatedness that had plagued earlier lexical decision studies (e.g., Azuma & Van Orden, 1997; Millis & Bution, 1989). On the other hand, using dictionaries to determine how many meanings or senses a word has may poorly correspond to native speakers' knowledge of that word (see Gernsbacher, 1984; Lin & Ahrens, 2005, 2010). This criticism provided a strong impetus to re-examine the effect of ambiguity using more psychologically valid measures – namely, subjective ratings of the number of meanings and their semantic relatedness. Contrary to the previous demonstrations of a polysemy advantage and a homonymy disadvantage (e.g., Beretta et al., 2005; Rodd et al., 2002), lexical decision studies that adopted this approach found that both homonymy and polysemy facilitate word recognition (Haro et al., 2017; Hino et al., 2006; Pexman et al., 2004). Follow-up studies showed that this is also the case when the task involves difficult (Hino et al., 2010) or pseudo-homophonic non-words (Haro & Ferré, 2018) that were argued to be necessary to increase semantic processing and produce the effect of relatedness (Armstrong & Plaut, 2008, 2016; Azuma & Van Orden, 1997; Rodd et al., 2002).

The reason for this conflicting evidence is far from clear (for a fairly recent discussion, see Eddington & Tokowicz, 2015). Studies of lexicographic polysemy (Armstrong & Plaut, 2008; Rodd et al., 2002) assume that lexical decisions are made based on initial semantic activation that is consistent and facilitatory for polysemes but inconsistent and inhibitory for homonyms. In contrast, studies of subjective polysemy (e.g., Haro et al., 2017; Hino et al., 2002) propose that lexical decisions are primarily made based on orthographic activation. Under this “semantic feedback” account (Hino & Lupker, 1996), activation at the semantic level does not underlie word-form identification but only facilitates the process through feedback to activation at the orthographic level. Since ambiguous words have multiple semantic representations, they produce stronger semantic activation (feedback), and are therefore recognised faster than unambiguous words. Critically, the assertion is that homonymy and polysemy do not differ in terms of representation, and that they should benefit word recognition (or hinder word comprehension) to the same extent.

This proposal, however, received weak support from other studies of subjective polysemy, in which sensicality judgements were made to homonyms and polysemes embedded in modifier-noun phrases (Klein & Murphy, 2001; Klepousniotou et al., 2008; Pylkkänen, Llinás, & Murphy, 2006; see also Brown, 2008). In the first study of this kind, Klein and Murphy (2001) found that responses to ambiguous targets (e.g., “liberal paper”) were slower when preceded by different-meaning (e.g., “wrapping paper”) than same-meaning primes (e.g., “daily paper”). Critically, the size of this interference effect did not differ between homonyms and polysemes, which led Klein and Murphy (2001) to conclude that both types of words have separate semantic representations that engage in competitive processes. In their follow-up MEG study, however, Pylkkänen et al. (2006) argued that this may not be true. The researchers modified the design, so that ambiguous targets (e.g., “liberal paper”) were preceded by related different-meaning (e.g., “lined paper”) or unrelated primes (e.g., “military post”). The results showed differences between homonyms and polysemes in the M350 component that is assumed to index semantic activation (e.g., Pylkkänen & Marantz, 2003). Homonymous targets elicited a later M350 latency after the related than the unrelated prime, indicating inhibition from the different, previously retrieved meaning. Polysemous targets, on the other hand, elicited an earlier M350 latency after the related prime, indicating facilitation from the different sense. This was taken as the first evidence, from sensicality judgments, against Klein and Murphy’s (2001) idea that the unrelated meanings and related senses of ambiguous words are represented and processed in the same manner.

Additional evidence came from Klepousniotou et al. (2008) who explored sensicality judgements to ambiguous nouns with low (e.g., “dinner/historical date”), moderate (e.g., “natural/confidential intelligence”), and high overlap in meaning (e.g., “bad/childhood dream”), as determined based on participants’ ratings of this property. The results generally showed that high-overlap words differed from moderate- and low-overlap words (for similar results, see Brown, 2008). When HF-meaning targets (e.g., “blind date”) followed LF-meaning primes (e.g., “expiration date”), a processing cost arose only for moderate- and low-overlap words. When LF-meaning targets (e.g., “historical date”) followed HF-meaning primes (e.g., “dinner date”), a processing cost also arose for high-overlap words, but it was significantly smaller than that for the other ambiguous words. Given this pattern of responses, Klepousniotou et al. (2008) claimed that previous studies of subjective polysemy, in particular Klein and Murphy (2001), failed to find a relatedness effect because they did not consider the degree to which the different senses of polysemes were related, but only ascertained that native speakers perceived them as related when contrasted with the different meanings of homonyms. Klepousniotou et al. (2008) proposed that it is the degree of semantic relatedness that plays a crucial role in both representation and processing, such that polysemes with loosely related senses may be more similar to homonyms than to polysemes with highly related senses.

Taken together, these studies of subjective polysemy make conflicting predictions regarding competition involved in the processing of multiple word senses. Lexical decision studies (Haro et al., 2017; Hino et al., 2006, Pexman et al., 2004) suggest that polysemy should slow word comprehension to the same extent as homonymy, although these studies typically attributed the ambiguity disadvantage to decision making, rather than semantic competition. In contrast, sensicality judgement studies, in particular Brown (2008) and Klepousniotou et al. (2008), suggest that the impact of polysemy may depend on the degree of sense relatedness, such that competition could be minimal or non-existent for polysemes with highly related senses but stronger for polysemes with loosely related senses. This prediction is compatible with the research into linguistic polysemy as reviewed in the next subsection.

5.1.3 Linguistic Polysemy

The view of ambiguity as a continuum (Brown, 2008; Klepousniotou et al., 2008) is further supported by studies that distinguished homonymy and different forms of polysemy by consulting their definitions within the theoretical linguistic literature (e.g., Ahrens, Chang, Chen, & Huang, 1998; Apresjan, 1974; Asher, 2011; Pustejovsky, 1995). In one such study, Klepousniotou and Baum (2007) drew a distinction between homonyms (e.g., “fan” as the device or a person), metaphorical polysemes (e.g., “fox” as the animal or a cunning person), and metonymic polysemes (e.g., “onion” as the plant or the vegetable) based on linguistic properties of these words. As discussed in Chapter 1, what differentiates polysemes from homonyms is the shared etymological derivation and conceptual overlap of their senses (e.g., Cruse, 1986; Lyons, 1977). What differentiates most metaphors from metonyms, on the other hand, is the nature of sense extension (e.g., Apresjan, 1974; Lakoff, 1987), whether the alternative sense is derived figuratively, rather than via productive rules of, for example, animal-for-meat (e.g., “rabbit”) and instrument-for-action alternations (e.g., “shovel”). Having classed their ambiguous words as such, Klepousniotou and Baum (2007) found that although lexical decisions benefited from both forms of polysemy, this effect was greater for metonyms than metaphors. In another study, Klepousniotou (2002) also showed this to be the case when lexical decisions to ambiguous words (e.g., “rabbit”) were cross-modally primed by sentences biasing a particular sense (e.g., “The hunter killed one” or “The chef made a stew”). Consistent with the research into lexicographic polysemy (e.g., Armstrong & Plaut, 2011; Beretta et al., 2005), these findings confirm that it is polysemy that facilitates language processing, but they also support the division of polysemy in the theoretical and computational linguistics literature (Apresjan, 1974; Geeraerts, 2010; Lopukhina et al., 2018).

Differences between homonyms, metaphors, and metonyms have been also found in two ERP studies (Klepousniotou et al., 2012; MacGregor et al.,

2015) that explored meaning/sense activation, as indexed by N400 priming effects in a lexical decision task. The studies used the same word stimuli but varied the inter-stimulus interval (ISI) to probe the time-course of activation (50 ms in Klepousniotou et al., 2012; 750 ms in MacGregor et al., 2015). The results revealed that the meanings of homonyms, mainly the dominant one, were activated at the short ISI but not at the long ISI. The senses of metonyms, on the other hand, were activated to the same extent regardless of the ISI (for similar behavioural results, see Williams, 1992). This was also the case for metaphors, although at the short ISI activation of the alternative sense was weaker in the left but not in the right hemisphere. Based on this pattern of results, MacGregor et al. (2015) proposed that the meanings of homonyms compete with each other, such that neither remains activated over a long period in the absence of context, whereas the senses of metaphors and metonyms cooperate in preventing such decay.

The studies of linguistic polysemy reviewed so far demonstrate that there are major differences in how homonyms and polysemes are processed and, by inference, represented in the mental lexicon. The finding that metonymy does not produce any interference typical of homonymy suggests that this form of ambiguity may involve only one semantic representation (Klepousniotou, 2002; Klepousniotou & Baum, 2007; MacGregor et al., 2015). Klepousniotou and Baum (2007), in particular, argued that such a representation must be rich in semantic information, or else one would struggle to explain why metonyms are recognised faster than their unambiguous counterparts that also have a single representation. As for metaphors, the evidence is less clear-cut. These words do not produce interference, nor do they facilitate processing as much as metonyms;

hence the tentative proposal is that metaphor lies somewhere between homonymy and metonymy (Klepousniotou, 2002; Klepousniotou & Baum, 2007).

The single-representation account for metonyms is consistent with several eye-tracking studies that specifically focussed on this type of ambiguity (Frazier & Rayner, 1990; Frisson & Frazier, 2005; Frisson & Pickering, 1999, 2007; McElree et al., 2006; Pickering & Frisson, 2001). Broadly speaking, this line of research revealed that, unlike homonyms, metonyms do not show a typical dominance effect; reading times for late-disambiguation sentences with metonyms in the dominant sense (e.g., “Apparently, the dinner wasn’t very enjoyable, tasting burned.”) are equivalent to those in the alternative sense (e.g., “Apparently, the dinner wasn’t very enjoyable, ending early.”). This suggests that readers do not commit to a specific interpretation as soon as metonyms are encountered but wait until further context provides definite support for such an interpretation, which has been taken as evidence that metonyms have only one representation that acts as a “gateway” for their multiple senses (for a review, see Frisson, 2009).

Note, however, that there is very little consensus on what information the representation of a metonym may actually encompass (for a detailed review, see Falkum & Vicente, 2015). One possibility is that the representation includes all features that are shared by the different word referents, often referred to as the “common core”, but generally maps onto the dominant or “basic” sense (e.g., Copestake & Briscoe, 1995; Klepousniotou et al., 2008; Pustejovsky, 1995). Under this “core representation” view, readers access the dominant sense (e.g., “rabbit” as the animal) directly from the representation whenever a metonym is encountered, but must derive the alternative sense (e.g., “rabbit” as the meat) via a productive rule (e.g., producer-for-product alternation) when that sense is intended. Another possibility is that the representation includes the common core in a more abstract and underspecified form (e.g., Caramazza & Grober, 1976; Frisson, & Pickering, 1999; Ruhl, 1989). Under this "underspecification" view, readers access a general meaning of a metonym when they encounter the word, and then use contextual information to home in on the dominant or the alternative sense. Critically, even though these two accounts disagree on what the core comprises and how readers disambiguate metonyms, they agree that the senses of these words share one representation.

It is important to mention that the single-representation account has been also proposed for so-called "irregular" polysemy (Brocher et al., 2016, 2018). Within the theoretical linguistic literature (Apresjan, 1974; Lehrer, 1990), this form of ambiguity is characterised by a unique pattern of sense extension (e.g., "wire" denoting a strand of metal or a listening device; "drone" denoting a male bee or an aircraft) that does not follow any productive rule typical of metonymy and cannot be found in any other words of the given language. Studies by Brocher et al. (2016, 2018) suggest that irregular polysemes are similar to metonymic polysemes, in that their alternative sense does not incur an extra processing cost either. Reading times for late-disambiguation sentences with irregular polysemes in the alternative sense were found to be equivalent to those with unambiguous control words (e.g., "When Mr. Jordon discovered the wire/bomb in the lamp, the FBI aborted the secret mission."), regardless of whether the polysemes had balanced or unbalanced sense frequencies. To accommodate their findings, Brocher et al. (2016) proposed that irregular polysemes have a single semantic representation that is divided into two components – namely, features that are shared by the different word referents and features that are not. Under this "shared features" view, readers first access an overall meaning of an irregular polyseme based on activation of the shared portion, and then disambiguate the word toward a specific sense, normally the more frequent one, based on activation of the unshared portion. The key assumption, then, is that the alternative senses of irregular polysemes are accessed directly from the mental lexicon, whilst those of metonyms are derived from context (e.g., Frisson, 2009) or via a productive rule (e.g., Klepousniotou et al., 2008).

Taken together, it appears that the impact of polysemy in comprehension may depend on the specific form of polysemy, as suggested by studies of subjective polysemy that varied the degree of sense relatedness (Brown, 2008; Klepousniotou et al., 2008). Competition between semantic representations in the race for activation should arise for metaphorical polysemes that seem to have separate or overlapping representations (e.g., Klepousniotou & Baum, 2007; Lopukhina et al., 2018), but not for metonymic and irregular polysemes that seem to have only one representation (e.g., Brocher et al., 2016, 2018; Frazier & Rayner, 1990; Frisson & Pickering, 1999; Klepousniotou et al., 2012).

5.1.4 Literature Synthesis

It is evident that the existing literature on ambiguity is convoluted and makes conflicting predictions on the role of meaning relatedness in semantic competition. A key factor seems to be one's operalisation of ambiguity. Studies of lexicographic polysemy (e.g., Armstrong & Plaut, 2011; Jager et al., 2016; Rodd et al., 2002) counted how many senses a given word had but did not determine their relationship. Studies of subjective (e.g., Klepousniotou et al., 2008; Pexman et al., 2004) and linguistic polysemy (e.g., Brocher et al., 2016; Klepousniotou, 2002; Rayner & Frazier, 1990), on the other hand, paid more attention to the nature of such relationship but did not control for sense count. Although most findings tend to suggest that the ambiguity disadvantage in word comprehension should be smaller for polysemes than homonyms, it is unclear, for example, whether this would lie in the existence of related senses in itself or their number.

In recent years, there have been a few attempts to bridge the lines of research, though with varying degrees of success. For instance, Jager and Cleland (2015) found that lexical decisions were faster to metaphors than metonyms, rather than vice versa (Klepousniotou, 2002; Klepousniotou & Baum, 2007), when the two were distinguished based on linguistic principles but then matched on the number of senses in the Wordsmyth Dictionary (Parks et al., 1998). The researchers acknowledged, however, that this discrepancy could have been due to more familiar sense extensions for their metaphorical (e.g., “parrot”, “snail”; animal for human trait) than their metonymic polysemes (e.g., “anchovy”, “chinchilla”; animal for product). In a different study, Haro and Ferré (2018) suggested that one’s approach to the selection of homonyms likely impacts on how these words are recognised relative to non-homonyms. Their lexical decision data showed an inhibitory effect when using dictionary entries, but the reverse when using subjective ratings, which led Haro and Ferré (2018) to conjecture that the inhibitory effect is a false positive specific to investigations relying on the lexicographic approach (e.g., Armstrong & Plaut, 2016; Rodd et al., 2002). Since dictionaries often list uncommon, archaic, and technical meanings but omit novel word usage and slang, what those investigations assumed to be a homonym/non-homonym may not have been processed as such by their participants. This finding, in particular, highlights the need to combine the different approaches to the study of semantic ambiguity, if the field is to progress and reach a consensus.

Unlike Haro and Ferré (2018), Rodd et al. (2012) sought to clarify the impact of ambiguity in word-form processing using an artificial language learning paradigm, in which adult participants learnt new, fictitious meanings for previously unambiguous words (e.g., “sip” as a small amount of computer data). Their third and final experiment, which involved demanding 4-day training, revealed shorter lexical decisions to trained than untrained words, indicating that new meanings had been sufficiently consolidated to affect word processing in a task that did not even require access to semantic knowledge. Interestingly, this effect was larger for trained words with new related than unrelated meanings, which corroborates the view that word recognition benefits from polysemy but not homonymy (e.g., Beretta et al., 2005; Klepousniotou & Baum, 2007; Rodd et al., 2002). In other words, the study demonstrated that, once integrated into the lexicon, new related and unrelated meanings influenced word recognition in the same way as polysemy and homonymy in existing words, suggesting that learning new meanings for familiar words in experimental settings closely mirrors the impact of ambiguity in natural language. It appears, then, that artificial lexicon paradigms provide a great avenue to test predictions and assumptions within the ambiguity literature. Foremost, they allow for accurate manipulation of the homonymous or the polysemous status of word stimuli, circumventing the issues involved in doing so based on dictionaries, subjective ratings, or linguistic criteria.

5.1.5 Present Experiment

The present experiment was designed to determine whether polysemy, as a whole, gives rise to competition at the semantic level, and, if so, whether to the same extent as homonymy. This was essential to develop a comprehensive account of the ambiguity disadvantage within this thesis and advance or place important constraints on existing PDP models (e.g., Borowsky & Masson, 1996; Kawamoto, 1993; Rodd et al., 2004), according to which all forms of semantic ambiguity should give rise to such competition. The question of whether the disadvantage effect arises for words with related senses, rather than unrelated meanings, was also of particular value to the broader literature on polysemy. Since the effect serves as an index of competition between multiple semantic representations, Experiment 5 could shed much-needed light on whether polysemes have separate (e.g., Hino et al., 2006; Klein & Murphy, 2001), separate but overlapping (e.g., Armstrong & Plaut, 2008; Rodd et al., 2004), or single representations (e.g., Frisson, 2009; Klepousniotou et al., 2008).

Thus, Experiment 5 contrasted the effects of homonymy and polysemy using an artificial language paradigm (as in Rodd et al., 2012). Participants learnt new meanings for otherwise unambiguous words that were either semantically related or unrelated to the existing meanings. The advantage of this method was two-fold. First, the experiment had rigorous control over meaning relatedness, combining the different approaches to the study of semantic ambiguity. Following linguistic criteria, new related meanings were designed to imitate irregular polysemy, whilst the unrelated counterparts were designed to imitate homonymy.

For the former, new meanings were loosely related to existing meanings through a single semantic feature and could not be derived via a productive rule. This manipulation of meaning relatedness was then confirmed based on subjective ratings. In addition, all words for which new meanings were created had a similar, small number of senses prior to the experiment. Second, the learning paradigm helped to hold constant other word properties that have been shown to modulate ambiguity effects. These include relative meaning frequency (Experiments 2-4 in this thesis; Armstrong et al., 2012; Duffy et al., 1988), word-form frequency (Hino & Lupker, 1996; Jager et al., 2016), concreteness (Jager & Cleland, 2016; Tokowicz & Kroll, 2007), as well as grammatical class (Grindrod et al., 2014; Mirman et al., 2010).

Training materials and procedures were largely based on those of Rodd et al. (2012, Experiment 3) who were successful in teaching participants a large number of new word meanings and demonstrated that their intensive 4-day training allowed those meanings to be sufficiently consolidated to influence online word recognition. This is also in line with studies of word learning which suggest that while a few exposures may be sufficient to learn new word forms, this knowledge is not normally integrated into the lexicon until after offline sleep-dependent consolidation has taken place (for a review, see Davis & Gaskell, 2009). This literature, in particular, motivated the decision to employ multi-day training that would allow new meanings to develop robust representations that could produce potential competition.

However, unlike Rodd et al. (2012), Experiment 5 examined the effects of homonymy and polysemy in word comprehension, as opposed to word recognition. The aim of the experiment was to compare the degree of semantic competition, or the disadvantage effect, resulting from the consolidation of new related and unrelated meanings. Thus, Experiment 5 involved a relatedness decision task in which trained words (e.g., “sip” denoting a small amount of computer data) were followed by words that related to the familiar/existing meaning (“sip-liquid”) or were unrelated (“sip-eel”). Participants’ responses to the same word pairs were compared before and after training. The task required disambiguation toward the existing, dominant meaning, and would as such indicate potential competition from the newly-learnt meaning. Overall, then, the present experiment focussed on the general distinction between homonymy and polysemy, without contrasting the linguistic forms of the latter (i.e., irregular, metaphorical, and metonymic polysemy). This aimed to confirm that the two differ in their influence on word comprehension in the first place, as predicted by most of the literature on meaning relatedness (e.g., Armstrong & Plaut, 2008; Frazier & Rayner, 1990; Klepousniotou, 2002; Rodd et al., 2002).

5.2 Method

5.2.1 Participants

Thirty students and members of staff [23 females, aged 20-35 (*M* = 26.6, *SD* = 5.3)] from the University of Leeds took part in the experiment in exchange for a £20 voucher. All participants were right-handed monolingual native speakers of British English with no known history of language-/vision-related difficulties or disorders. The experiment received ethical approval from the School of Psychology, University of Leeds Ethics Committee.

5.2.2 Materials

5.2.2.1 New Word Meanings

Thirty-two "trained" words and short paragraphs (86-94 words) describing their new related meanings (e.g., "sip" denoting a very small amount of hacked computer data) were taken from Rodd et al. (2012)[26]. The paragraphs used each word in its new meaning five times, such that each instance provided a different piece of information about the new word referent (e.g., one sentence explained what a sip was, whereas another mentioned that extracting data in sips prevents hackers from getting caught). Most of the new meanings referred to recent inventions, colloquial and scientific terms, or social phenomena, and they were related to the existing meanings with respect to function (e.g., "bone" as the core of a star; n = 5), physical properties (e.g., "foam" as a type of nuclear waste; n = 12), being a specific variant of a more general meaning (e.g., "crew" as a group of musicians; n = 7), or the imagery that the word elicited (e.g., "hive" as a busy household; n = 8)[27]. Therefore, as in existing irregular polysemes, new meanings were related to original meanings through a single feature and could not be derived via a productive rule (e.g., animal-for-meat or part-for-whole relations) as the relationship between the meanings was unique to each word and

[26] The word "slim" in Rodd et al.'s (2012) stimulus list was changed to "mouse" so that all trained words had noun/noun-verb interpretations.

[27] As the experiment was not explicitly designed to explore the type of the relationship between the new and the existing meaning (e.g., physical properties vs. function), future studies will need to establish whether there could be an impact on learning performance based on the way new meanings are related.

unpredictable. New unrelated meanings were created by swapping the paragraphs across pairs of trained words (see the paragraphs in Appendix 3) to minimise any overlap between the related and unrelated meanings for each word. This manipulation of meaning relatedness was shown to be highly successful in a previous study using these materials (Maciejewski, 2014)[28]. Two versions of paragraphs were created, so that each contained 16 words with new related meanings and 16 words with new unrelated meanings. The related meanings in Version 1 were presented as unrelated in Version 2, and vice versa. All participants were pseudo-randomly assigned to learn from either version.

5.2.2.2 Relatedness Decisions

Each trained word served as a prime in the relatedness decision task assessing the comprehension of existing meanings (for the rationale behind the task, see 1.5 Aims of the Thesis in Chapter 1). To investigate potential practice or session effects on task performance, the stimulus list also included 32 untrained control primes that did not feature in any of the training materials. All trained and untrained prime words had noun or noun-verb interpretations and only one meaning in the Wordsmyth Dictionary (Parks et al., 1998). Although both trained and untrained words had a few related senses, neither exhibited patterns of sense extension typical of metaphorical (e.g., animal-for-human-characteristic relations) or metonymic polysemy (e.g., animal-for-meat relations). The two types

---

[28] Maciejewski (2014) reported that participants, who had just finished the 4-day training, rated new meanings in the related condition (M = 4.5, SD = 0.6) as more related to existing meanings than new meanings in the unrelated condition (M = 1.9, SD = 0.6). Ratings were made on a scale from 1 (“highly unrelated”) to 7 (“highly related”).

of primes were also statistically comparable (all *t*s < 1) with respect to 13 lexical and semantic variables, such as form frequency and the number of related word senses (see stimulus properties in Table 5.1 below).

Table 5.1: Experiment 5: Properties of the trained and untrained prime words.

Variable	Trained primes	Untrained primes
Letters	4.4 (1.0)	4.4 (1.1)
Phonemes	3.5 (0.8)	3.4 (0.9)
Syllables	1.1 (0.3)	1.2 (0.4)
Raw frequency	17.1 (20.0)	17.2 (15.1)
Log frequency	1.0 (0.5)	1.1 (0.4)
Orthographic neighbours	7.1 (5.4)	7.1 (6.8)
Log bigram frequency	2.8 (0.5)	2.7 (0.5)
Subjective familiarity	5.0 (0.5)	5.1 (0.5)
Senses	4.8 (2.0)	4.7 (2.1)
Semantic diversity	1.6 (0.2)	1.5 (0.2)
Concreteness	5.5 (0.7)	5.6 (0.8)
Imageability	5.6 (0.7)	5.7 (0.7)
Age of acquisition	6.2 (1.4)	6.2 (1.9)

Note. Standard deviations are given in the parentheses. Each prime set contained 32 items. Word-form frequency, bigram frequency, and number of orthographic neighbours come from the BNC (2007). “Senses” refers to the number of related word senses in the Wordsmyth Dictionary (Parks et al., 1998). Semantic diversity values come from Hoffman et al. (2013). Concreteness, imageability, and familiarity ratings come from the MRC Psycholinguistic Database (Coltheart, 1981). Age-of-acquisition ratings come from Kuperman et al. (2012).

Each prime word was paired with two target words – one related to the existing but not the new meaning (e.g., “sip-juice”) and the other unrelated to either meaning of the prime (e.g., “sip-golf”). Most of the targets related to primes through physical properties (e.g., “silk-satin”), category membership (e.g., “ant-insect”), and synonymy (e.g., “heap-mound”). All targets were nouns with a single entry in the Wordsmyth Dictionary (Parks et al., 1998), and their properties (see Table 5.2 below) were matched, at the group level, between the conditions involving trained and untrained primes (all *F*s < 1). Prior to the experiment, 30 monolingual native speakers of British English [15 females; aged 18-38 (*M* = 29.9, *SD* = 5.7)] rated prime-target relatedness on a 7-point scale (where 1 denoted “highly unrelated” and 7 denoted “highly related”). This stimulus pre-test confirmed that the related/unrelated pairs were considered as such, and that the trained (related pairs: *M* = 6.2, *SD* = 0.3; unrelated pairs: *M* = 1.9, *SD* = 0.4) and untrained prime words (related pairs: *M* = 6.2, *SD* = 0.3; unrelated pairs: *M* = 1.9, *SD* = 0.4) did not significantly differ with respect to the degree of semantic relatedness/unrelatedness (both *t*s < 1). All word pairs are given in Appendix 4.

Table 5.2: Experiment 5: Properties of the related and unrelated target words.

Variable	Trained primes		Untrained primes	
	Related targets	Unrelated targets	Related targets	Unrelated targets
Prime-target relatedness	6.2 (0.3)	1.9 (0.4)	6.2 (0.3)	1.9 (0.4)
Letters	5.0 (1.2)	5.0 (1.1)	4.9 (1.2)	5.0 (1.2)
Phonemes	4.1 (1.3)	4.2 (1.2)	4.3 (1.3)	4.2 (1.3)
Syllables	1.6 (0.7)	1.5 (0.6)	1.6 (0.7)	1.5 (0.7)
Raw frequency	21.2 (19.8)	21.4 (16.9)	21.1 (21.9)	21.4 (20.3)
Log frequency	1.1 (0.5)	1.2 (0.4)	1.2 (0.4)	1.2 (0.4)
Orthographic neighbors	3.5 (3.8)	3.6 (3.7)	3.7 (4.7)	3.6 (4.2)
Log bigram frequency	2.5 (0.4)	2.6 (0.5)	2.6 (0.6)	2.7 (0.5)
Subjective familiarity	5.3 (0.6)	5.2 (0.6)	5.1 (0.6)	5.2 (0.6)
Senses	3.8 (1.7)	3.8 (2.0)	3.7 (2.4)	3.8 (2.2)
Semantic diversity	1.5 (0.2)	1.5 (0.2)	1.5 (0.2)	1.5 (0.2)
Concreteness	5.7 (0.5)	5.7 (0.8)	5.8 (0.6)	5.8 (0.5)
Imageability	5.6 (0.6)	5.7 (0.8)	5.8 (0.4)	5.7 (0.7)
Age of acquisition	6.1 (1.8)	6.1 (1.9)	6.0 (1.7)	6.0 (1.9)

Note. Standard deviations are given in the parentheses. Each target set contained 32 items. Information on the different variables is given in the note for Table 5.1.

5.2.2.3 Worksheets

All participants completed three online worksheets, adapted from Rodd et al. (2012), on three consecutive days (Days 2-4) to help them further consolidate new meanings before their final testing session on Day 5. On Day 1, participants first read the paragraphs describing the new meanings of words. On Day 2, there were two exercises within Worksheet 1. Worksheet 1a involved selecting trained words from a drop-down menu and matching them to brief definitions of their new meanings, while Worksheet 1b involved answering one open-ended question about each new word referent. On Day 3, Worksheet 2 involved writing a new example sentence for each trained word that was compatible with its new meaning. On Day 4, Worksheet 3 involved writing a coherent story using all trained words in their new-meaning context. There was no word-count limit, and participants could write in any style and on any subject. However, they had to use each of the words at least once. The trained words were presented randomly in Worksheets 1a and 1b but alphabetically in Worksheets 2 and 3. The worksheets were designed and administered using Qualtrics (http://qualtrics.com/). All participants had access and could use the paragraphs when completing each of these worksheets.

5.2.3 Procedure

The experiment (for an overview, see Figure 5.1 below) took place over five consecutive days and lasted for four hours in total. Following Rodd et al. (2012, Experiment 3), the experiment consisted of an initial lab-based training session on Day 1, three home-based training sessions involving the online worksheets on Days 2-4, and a final lab-based testing session on Day 5. On Day

1, participants completed the pre-training relatedness decision task and then read paragraphs describing new word meanings. Over the next three days (Days 2-4), participants completed the online worksheets. On Day 5, they came back to the lab to complete the same relatedness decision task (using the same stimuli as on Day 1), followed by a recall task assessing their explicit memory for new meanings. Each participant completed the two lab-based sessions at a similar time of the day (+/- 2 hours), exactly five days apart. All the lab-based tasks were programmed in EPrime 2.0 (Schneider et al., 2012).

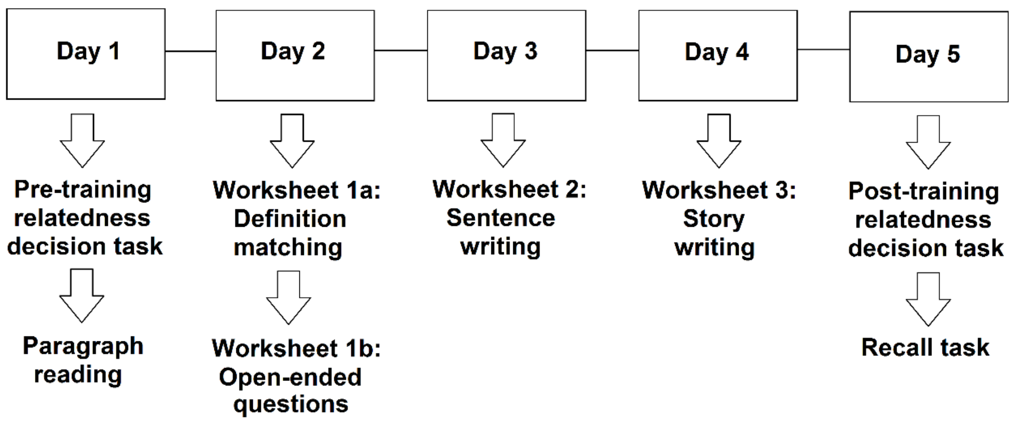

Figure 5.2: Experiment 5: Overview of the experiment.

5.2.3.1 Relatedness Decisions

As in Experiments 2-4, the task in Experiment 5 was to decide whether the prime and the target were related in meaning by pressing keyboard buttons (L for "yes" with their dominant, right hand, A for "no" with their left hand). On both testing sessions (Days 1 & 5), the task began with 20 practice trials with feedback on response accuracy and latency. The stimuli were divided into two blocks whose order was counterbalanced across participants. One block included 64 related pairs involving 32 trained and 32 untrained primes and 64 unrelated pairs serving as fillers (which were excluded from analyses). The other block included 64 unrelated pairs involving 32 trained and 32 untrained primes and 64 related fillers. None of the primes appeared more than once within the same block, and the fillers did not include any of the words used in the experimental stimulus list. The order of trials in each block was pseudo-randomised, such that no more than three related/unrelated trials appeared consecutively. There were two one-minute breaks – one after the practice block and one after the first experimental block. Each experimental block began with eight fillers (excluded from analyses) to help participants get back to the habit of quick responding following a break.

As in Experiment 2, trials began with a 500 ms fixation cross. After a delay of 100 ms, primes were presented for 200 ms followed by targets presented for 500 ms, with a delay of 50 ms in between. Participants were allowed an additional 1500 ms to respond. As soon as a response was made or at the end of the 1500 ms, there was a 100 ms delay before the next trial began. Participants could make a response 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). Response speed and accuracy were equally emphasised in the instructions, and participants were told what constituted semantic relatedness and given examples. The instructions on Day 5 were the same as those on Day 1 and did not mention anything about new word meanings.

5.2.3.2 Paragraphs

Following the pre-training relatedness decision task on Day 1, participants read paragraphs describing new meanings. The paragraphs were presented on a computer screen, one at a time and in randomised order. Participants pressed the spacebar to indicate when they had finished reading each paragraph. To ensure that they read the text slowly and carefully, 500 ms after having pressed the spacebar each paragraph was followed by a yes-no question relating to a specific feature of the new word referent (e.g., “Can only hackers extract sips?”). Once participants answered the question (by pressing the L button for “yes” or the A button for “no”), the next paragraph appeared after 1000 ms feedback on response accuracy and a delay of 100 ms. There was an equal number of “yes” and “no” responses in the task. All participants had as much time as they needed to read the paragraphs and answer the questions.

5.2.3.3 Worksheets

At the end of Day 1, participants received a paper booklet containing all the paragraphs and were instructed to use it as a companion for the worksheets. The order of worksheets was the same for all participants (i.e., Worksheet 1 on Day 2, Worksheet 2 on Day 3, Worksheet 3 on Day 4). They received access to a given worksheet at 8 a.m. on each day and had to complete it by midnight of that day. All participants completed the worksheets within this timescale.

5.2.3.4 Recall

On Day 5, participants came back to the lab and first performed the same relatedness decision task as on Day 1. They then completed a recall task in which they recalled and typed a maximum of nine features/properties that were true of the new word referents only. Participants had as much time as they needed to complete this task but could not use the companion booklet. They typed in “nothing” if they could not recall any information and pressed the ALT button to move to another word which appeared after a delay of 100 ms. The trained words were presented one a time and in randomised order.

5.3 Results

5.3.1 Worksheets

The first aim was to analyse learning performance, both during and after the training on new meanings. Worksheet results are summarised in Table 5.3 below. For Worksheet 1a (definition matching), one mark was assigned for each word that was correctly matched to the definition of its new meaning. For Worksheet 1b (open-ended questions), one mark was assigned for each correctly answered question about a new word referent. For Worksheets 2 (sentence writing) and 3 (story writing), participants received one mark for each trained word in the new-meaning context, regardless of how many times that word was used. The analysis of Worksheet 2 results excluded 10 participants who provided definitions of the new word referents rather than their own example sentences. The analysis of Worksheet 3 results excluded 3.2% of responses that lacked in detail and may have instantiated existing meanings.

The researcher (GM) first attempted to analyse responses using logit mixed-effects modelling, but this was not warranted – no random effects were significant (i.e., the number of correct responses did not substantially vary across subjects or items). A set of by-subjects (F_1) and by-items (F_2) ANOVAs with the factors of Meaning Type (new related meaning, new unrelated meaning) and Version (1, 2) was used instead. There were no effects of Version in any of the tasks. Thus, throughout the chapter, effects involving Version are not reported as the sole purpose of this factor was to account for the potential influence of counter-balancing (Pollatsek & Well, 1995). As expected, there were no effects of Meaning Type on either of the worksheets (all Fs < 1). The overall performance was at ceiling, most likely because participants could use the paragraphs when completing the worksheets. This confirms that the online training provided an opportunity to further consolidate both new related and unrelated meanings.

Table 5.3: Experiment 5: Mean percentages of correct responses for the online worksheets.

Meaning Type	Worksheet 1a	Worksheet 1b	Worksheet 2	Worksheet 3
New related meaning	99.4 (2.5)	99.1 (2.2)	98.7 (3.5)	98.7 (3.5)
New unrelated meaning	98.3 (3.6)	99.6 (2.3)	98.3 (3.7)	98.6 (2.8)

Note. Standard deviations are given in the parentheses.

5.3.2 Recall

For the recall task, participants received one mark for each of the five properties of the new word referents that were stated in the paragraphs. As in Rodd et al. (2012), two separate analyses were conducted – one for the number of “correct responses” (i.e., responses to trained words for which at least one property was correctly recalled) and the other for the number of correctly recalled properties for correct responses only (i.e., a maximum of five properties per new word referent). Overall, participants' recall performance was fairly good - the percentage of correct responses ranged (across participants) from 50% to 100% (*M* = 89.9%, *SD* = 15.1). Most of the incorrect responses were null ("nothing") responses (64%), with the remaining responses being "transfer errors" (i.e., recalling a property of a different new word referent).

Numbers of correct responses were analysed using a logit Meaning Type × Version mixed-effects model that included a significant random intercept for subjects. The analysis [$\chi^2(1) = 33.1$ $p < .001$; marginal $R^2 = .07$, conditional $R^2 = .55$] showed that the percentages of correct responses were significantly higher for the words with new related (*M* = 94.4%, *SD* = 12.3) than unrelated meanings (*M* = 84.4%, *SD* = 19.0).

Numbers of correctly recalled properties for correct responses were analysed using a linear Meaning Type × Version mixed-effects model that included significant intercepts for subjects and items and a random slope for Meaning Type across items. The model [$\chi^2(1) = 0.8$, $p = .37$; marginal $R^2 = .01$, conditional $R^2 = .38$] showed that Meaning Type did not influence the number of correctly recalled properties (related meaning: *M* = 3.7, *SD* = 0.6; unrelated meaning: *M* = 3.8, *SD* = 0.6).

5.3.3 Relatedness Decisions

The main aim of this experiment was to establish the influence of learning new meanings on the processing of existing meanings. Two of the 30 participants were removed from all analyses of the relatedness decision task – one due to a rather small number of correct responses in the recall task (50.0%) and the other due to slow responses across all trials (*M* = 870.0 ms, *SD* = 129.0). Analyses of both accuracy and latency excluded trials involving trained primes for which participants could not recall any property of their new word referents (4.5% of all responses). This was necessary to ensure that the training effect was examined only for words with truly consolidated meanings. For RTs, analyses also excluded errors (4.3% of the remaining responses) and outliers (defined as two standard deviations above/below a participant's mean per condition; 4.1%). RTs were log-transformed to normalise the distribution of residuals, hence below are reported back-transformed means of RTs and their CIs that were estimated from mixed-effects models.

The first set of analyses combined the trained prime words across the levels of Meaning Type (new related/unrelated meaning) and compared them to the untrained prime words. The rationale was that since this experiment involved unequal numbers of primes (16 words with new related/unrelated meanings and 32 untrained words), any direct comparisons across the three types of primes would be biased. Accuracy and latency data were analysed using logit/linear mixed-effects models with the factors of Session (pre-training, post-training), Prime (trained, untrained), Target (related, unrelated), and Block (1, 2)[29]. Terms involving Version were excluded due to model non-convergence. All models included random intercepts for subjects and items. The random slope for the Session × Target interaction across subjects and the random slope for Session

---

[29] There were no effects of Block in this experiment, neither in the accuracy nor the latency data.

across items were significant and included in the latency but not the accuracy model. Post hoc tests (with the Bonferroni correction for multiple comparisons) were used to explore significant interactions.

Mean error rates (%) for the trained and untrained primes are illustrated in Figure 5.2 below. There were no significant main effects in the accuracy model (marginal R^2 = .02, conditional R^2 = .36). Analyses revealed a Session × Target interaction [$\chi^2(1) = 6.7$, $p < .01$] that was due to a significant increase in post-training error rates for unrelated (M_{pre} = 3.3%, *SD* = 3.0; M_{post} = 4.7%, *SD* = 4.7; $p < .05$) but not related targets (M_{pre} = 5.3%, *SD* = 5.0; M_{post} = 4.2%, *SD* = 3.8; $p = .27$). There was also a significant Session × Prime × Target interaction [$\chi^2(1) = 3.9$, $p < .05$]. Post hoc tests indicated that this interaction concerned the trained primes only. Following the training, error rates for these words were lower on trials involving related targets (M_{pre} = 6.8%, *SD* = 7.1; M_{post} = 3.9%, *SD* = 3.8; $p < .05$), but not on trials involving the unrelated counterparts (M_{pre} = 3.4%, *SD* = 4.7; M_{post} = 5.2%, *SD* = 6.2; $p = .16$).

Mean RTs (ms) for the trained and untrained primes are illustrated in Figure 5.3 below. The latency model (marginal R^2 = .09, conditional R^2 = .54) revealed a significant main effect of Target [$\chi^2(1) = 25.3$, $p < .001$], with slower responses to unrelated (*M* = 632.9 ms, 95% CIs: 598.8, 668.7) than related targets (*M* = 571.5 ms, 95% CIs: 545.9, 598.3). Responses were also slower on the post-training (*M* = 613.1 ms, 95% CIs: 581.0, 646.8) than the pre-training session (*M* = 589.8 ms, 95% CIs: 562.6, 618.4), although this effect of Session only approached the significance threshold [$\chi^2(1) = 3.3$, $p = .07$]. There was a significant main effect of Prime [$\chi^2(1) = 27.3$, $p < .001$], with slower responses to the trained primes (*M* = 618.7 ms, 95% CIs: 589.4, 649.5) than the untrained counterparts (M = 584.4 ms, 95% CIs: 556.7, 613.6). Finally, the model also revealed a significant Session × Prime interaction [$\chi^2(1) = 31.6$, $p < .001$]. Post hoc tests showed a significant increase in post-training RTs for the trained (M_{pre} = 598.4 ms, 95% CIs: 570.0, 628.4; M_{post} = 639.7 ms, 95% CIs: 605.6, 675.6; $p < .001$) but not untrained primes (M_{pre} = 581.3 ms, 95% CIs: 553.7, 610.4; M_{post} = 587.5 ms, 95% CIs: 556.3, 620.6; $p = 1$). No other effects approached the significance threshold.

These analyses demonstrate that having learnt new meanings slowed participants' responses to previously unambiguous words. To investigate the role of the semantic relatedness between the existing and the new meaning, the second set of analyses excluded the untrained primes and directly compared the two types of trained primes. Response-accuracy and response-latency models included the same fixed effects as those in the models above, except that Prime was replaced with Meaning Type (new related vs. new unrelated). With respect to random effects, both models included random intercepts for subjects and items. The latency model also included random slopes for the Session × Target and Meaning Type × Target interactions across subjects as well as a random slope for Session across items.

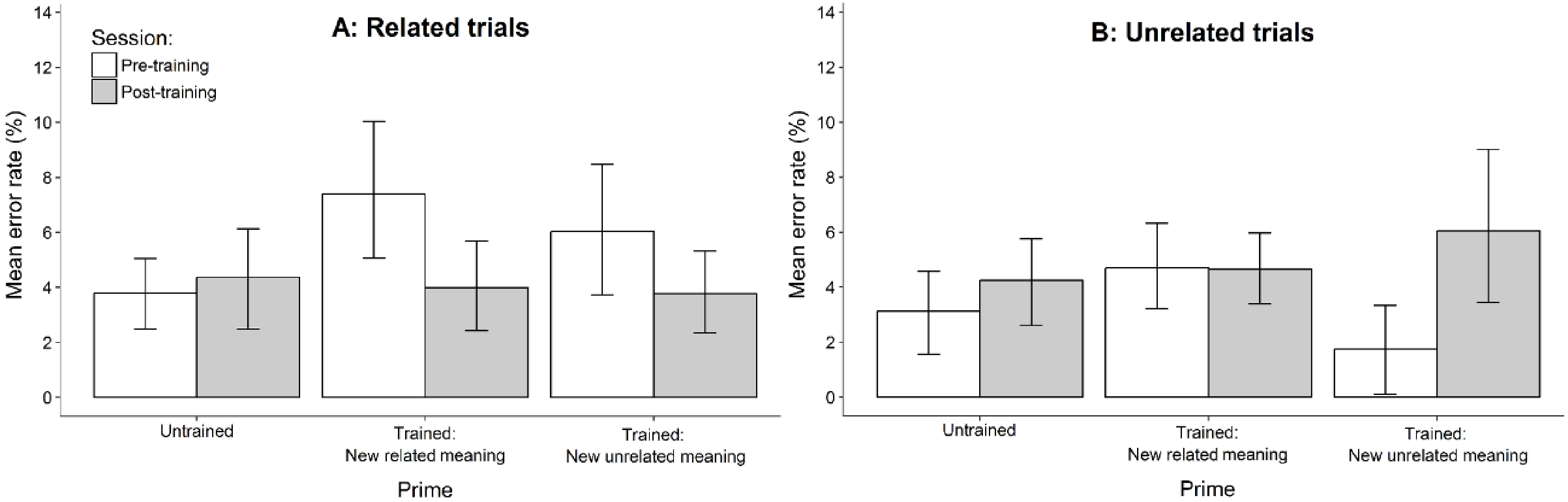

Figure 5.3: Experiment 5: Subject means of error rates for related (Panel A) and unrelated targets (Panel B). Error bars show 95% confidence intervals adjusted to remove between-subjects variance.

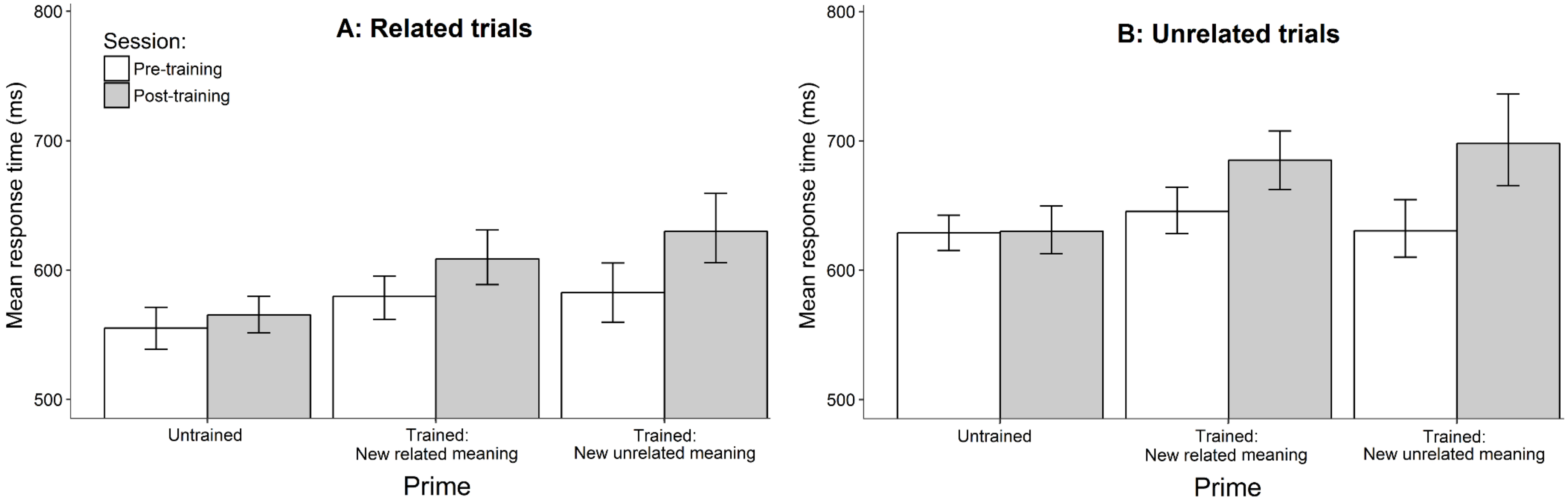

Figure 5.4: Experiment 5: Subject means of untransformed RT for related (Panel A) and unrelated targets (Panel B). Error bars show 95% confidence intervals adjusted to remove between-subjects variance.

The accuracy model (marginal R^2 = .06, conditional R^2 = .45) revealed only a significant Session × Target interaction [$\chi^2(1) = 11.4$, $p < .001$]. Post hoc tests indicated that following the training, error rates decreased for related targets (M_{pre} = 6.8%, SD = 7.1; M_{post} = 3.9%, SD = 3.8; $p < .05$), but increased for unrelated targets (M_{pre} = 3.4%, SD = 4.7; M_{post} = 5.2%, SD = 6.2; $p < .05$).

In contrast, the latency model (marginal R^2 = .07, conditional R^2 = .54) revealed a significant main effect of Target [$\chi^2(1) = 15.0$, $p < .001$], with faster relatedness decisions to related (M = 591.0 ms, 95% CIs: 562.5, 620.9) than unrelated targets (M = 648.9 ms, 95% CIs: 610.8, 689.5). Responses were also significantly slower on the post-training (M = 640.5 ms, 95% CIs: 605.3, 677.6) than the pre-training session (M = 598.7 ms, 95% CIs: 569.6, 629.4), and this effect of Session [$\chi^2(1) = 8.5$, $p < .01$] interacted with Meaning Type [$\chi^2(1) = 5.6$, $p < .05$]. Post hoc tests indicated that the simple effect of Session was significant for both the words with new unrelated (M_{pre} = 595.0 ms, 95% CIs: 565.3, 626.3; M_{post} = 645.1 ms, 95% CIs: 609.1, 683.3; $p < .001$) and related meanings (M_{pre} = 602.4 ms, 95% CIs: 573.2, 633.3; M_{post} = 635.9 ms, 95% CIs: 600.9, 672.8; $p < .01$), but it was significantly greater for the former. All other effects did not approach the significance threshold.

5.4 Discussion

Experiment 5 showed that participants learnt many of the new meanings over the course of the intensive training, though their ability to recall them was superior for related than unrelated meanings. Consistent with Rodd et al. (2012), relatedness in meaning benefitted the likelihood of access to the semantic representations for newly-acquired meanings but not the amount of information within the representations. Participants recalled as many semantic features for related word referents as they did for the unrelated counterparts, whenever they correctly recalled any information about new meanings. This suggests that the overlap in features between new and existing meanings acts as a cue during the learning and/or retrieval of new meanings, but it does not determine the robustness or richness of their semantic representations. With regard to the impact of consolidation, Experiment 5 showed that learning new meanings slowed participants' comprehension of existing meanings. This effect, which was observed for both related and unrelated prime-target word pairs, was significantly smaller for meanings that were related to existing meanings than for the unrelated counterparts. Taken together, the results demonstrated that relatedness in meaning influences the learning of new word meanings and their subsequent impact on semantic processing.

5.4.1 Meaning Relatedness

The present experiment revealed that learning new meanings influenced the processing of previously unambiguous words in a semantically engaging online task, indicating that the meanings had been successfully "lexicalised" (Gaskell & Dumay, 2003) or "engaged" within the mental lexicon (Leach & Samuel, 2007). As expected, consolidation of new meanings hindered the comprehension of existing meanings, mirroring the processing disadvantage observed in studies using existing ambiguous words (e.g., Experiments 2-4 in this thesis; Gottlob et al., 1999; Hoffman & Woollams, 2015). This finding lends further support to the semantic competition account that comes from PDP models of word processing (e.g., Borowsky & Masson, 1994; Kawamoto et al., 1994; Rodd et al., 2004). In line with this account, it seems that, once integrated into the mental lexicon through extensive training and offline consolidation, new meanings began to compete with well-established meanings in the race for activation. The experiment further delineated this competition by demonstrating that it is modulated by the degree of semantic relatedness between the new and the existing meaning, such that the greater the relatedness, the smaller the competition.

This effect of meaning relatedness has important implications for PDP models that recognise the role of that property in ambiguity representation and processing, such as the ones proposed by Armstrong and Plaut (2008) and Rodd et al. (2004). While both models assume competition for homonyms, they disagree on competition for polysemes. Consistent with the present findings, the model by Rodd et al. (2004) predicts that competition for polysemes should be smaller than that for homonyms due to overlap in semantic features. Rodd et al. (2004) suggest that polysemes have separate but overlapping representations, and that this should result in reduced competition that involves only those features that are unique to the different senses.

In contrast, the model by Armstrong and Plaut (2008) predicts that polysemy would not slow word comprehension at all. According to their model, polysemes also have separate overlapping semantic representations, but any competition between the representations is cancelled out by a processing benefit at the earlier stages of word processing. As discussed in the introduction to Experiment 5, a number of lexical decision studies showed that polysemy facilitates visual word recognition (e.g., Beretta et al., 2005; Klepousniotou & Baum, 2007; Rodd et al., 2002). This led Armstrong and Plaut (2008) to predict that the polysemy advantage during orthographic processing may be equal to the polysemy disadvantage during semantic processing, such that the former eliminates the latter in tasks that require both processing stages to be completed (e.g., relatedness decisions). However, while Rodd et al.'s (2012) lexical decision task showed that the learning of new related meanings can indeed facilitate word recognition, the current findings, from a semantically engaging task involving the same word stimuli, reveal that the learning still slows word comprehension. It appears that the polysemy advantage during orthographic processing does not entirely cancel out the polysemy disadvantage during semantic processing. Therefore, even at the relatively early stages of meaning consolidation, new meanings of irregular polysemes produce some degree of competition when the task requires meaning selection.

The effect of meaning relatedness reported in this experiment presents another major challenge for the decision-making account. Although Pexman et al. (2004) also found that the processing disadvantage (on related trials) was smaller for polysemes than homonyms, they explained this result in terms of response-conflict resolution, rather than semantic processing. Consistent with Experiments 2-4, however, this experiment shows that not only did the training slow relatedness decisions on related trials that are assumed to involve response conflict, but also on unrelated trials where the new and the existing meaning triggered the same response. If the training effect were due to decision making during response selection, one would not expect to find it on unrelated trials.

Pexman et al.'s (2004) account, therefore, fails to explain why newly-acquired meanings would hinder the ability to comprehend well-established meanings.

It is also improbable that the training effect was due to a "checking" process (Hino et al., 2006), or a task strategy whereby participants took additional time to ensure that the targets were not related to new meanings. The results indicate that the effect was smaller for new related meanings, supporting the proposal that the competition involved in processing existing ambiguous words may be modulated by the degree of overlap in semantic features (Brown, 2008; Klepousniotou et al., 2008; Rodd et al., 2004). The fact that the training effect, like the ambiguity effect in natural language, is sensitive to relatedness in meaning strongly suggests that the slowing lies in semantic, rather than task-specific decision-making, processes.

Experiment 5 is also relevant to the broader literature on ambiguity that has to date produced mixed evidence on the role of meaning relatedness. It shows, for the first time, that although both new related and unrelated meanings entail competition, the former do so to a lesser extent. This effect corroborates the emerging view that homonyms and polysemes differ in representation and processing, as consistently suggested by studies of lexicographic (e.g., Beretta et al., 2005; Rodd et al., 2004) and linguistic polysemy (e.g., Frazier & Rayner, 1990; MacGregor et al., 2015; Klepousniotou, 2002), and less consistently by studies of subjective polysemy (e.g., Brown, 2008; Klepousniotou et al., 2008; Pylkkänen et al., 2006). Note, however, that the effect is incompatible with recent suggestions that irregular polysemes may have a single representation that comprises both shared and unshared semantic features of the different word referents (Brocher et al., 2016, 2018). Under this account, new related meanings in Experiment 5, which imitated irregular polysemy in natural language, should not produce competition at all. Their unique, or unshared, semantic features should be integrated into the existing representation for the previously unambiguous word, rather than into a new representation. Since learning new related meanings did produce competition, the present findings are more compatible with the idea that the loosely and idiosyncratically related senses of irregular polysemes have separate representations that overlap in semantic space (Brown, 2008; Klepousniotou et al., 2008; Rabagliati & Snedeker, 2013; Rodd et al., 2004), but still engage in competition for semantic activation.

5.4.2 Vocabulary Acquisition

The work reported in this chapter helps to advance our understanding of vocabulary acquisition in adults. Although there have been multiple studies on learning new words (for a detailed review, see Davis & Gaskell, 2009), little is known about adults' ability to learn new meanings for words they already know – an important prerequisite for skilled language use. Experiment 5 extends this work on word learning into the semantic domain, delineating how new meanings are integrated into existing lexical-semantic representations, and how they affect access to those representations. In particular, it demonstrates that while it is fairly easy to learn new (related) meanings, this information comes at a cost in the form of semantic competition.

Such competition bears a striking resemblance to lexical competition reported in studies of word learning (e.g., Bowers, Davis, & Hanley, 2005; Gaskell & Dumay, 2003; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). The general finding of these studies is that consolidation of new word forms (e.g., “cathedruke”) slows the recognition of known neighbours (e.g., “cathedral”), in either the spoken or the written modality. Although there are differences between learning new meanings for familiar words and learning new words, it seems that integration of both types of information hinders processing because of the way lexical-semantic representations are formed and accessed. The implication is that, just like lexical competition has served as an index of consolidation of new word forms, semantic competition, documented in this chapter, can serve as an index of consolidation of new word meanings. Thus, this work provides the field with a novel paradigm to address crucial questions about meaning consolidation. These include the role of overnight sleep, learning performance across the lifespan, the nature of training (e.g., naturalistic and semantically diverse context vs. dictionary definitions), and the time-course of meaning consolidation. The latter question is of particular value since recent studies (Fang & Perfetti, 2017, 2018; Fang, Perfetti, & Stafura, 2017) have suggested that new and existing meanings of words may engage in competitive processes even shortly after the learning period.

It should be noted that the implications of the current work are restricted to learning new meanings that are loosely related or unrelated to existing meanings. The findings make no prediction with respect to learning new senses that follow the rules of sense extension characteristic of metonymy, such as instrument for action (e.g., “shovel”) and container for contents alternations (e.g., “bottle”). Studies have shown that adults (Clark & Gerrig, 1983; Frisson & Pickering, 2007; McElree et al., 2006; Murphy, 2006) and even four-year old children (Srinivasan, Al-Mughairy, Foushee, & Barner, 2017; Srinivasan, Berner,

& Rabagliati, 2018; Srinivasan & Snedeker, 2011, 2014) have no apparent difficulty working these senses out. Furthermore, there is notable evidence that metonyms, whose senses share a large number of semantic features, have only one representation, and may thus escape competition (e.g., Frazier & Rayner, 1990; Frisson & Pickering, 1999; Klepousniotou, 2002; Lopukhina et al., 2018). It is plausible, then, that new senses of metonyms do not require explicit learning or integration into the lexicon but can be derived on-line via a productive rule.

5.4.3 Conclusions

In summary, the finding that consolidation of new meanings for known words hinders the comprehension of their existing meanings has important implications for the study of the ambiguity disadvantage. In particular, it lends support to the postulate of semantic competition in PDP models of ambiguity processing, especially those that predict at least some competition for polysemes (Rodd et al., 2004). This competition could be further modulated by the degree of sense relatedness, such that competition could be minimal or even non-existent for the highly related senses of metonyms but stronger for the loosely related senses of irregular polysemes (as the ones used here). The present experiment also adds a novel type of evidence for the differential representation and processing of homonymy and polysemy. It shows the relatedness effect, both in learning and processing, using an artificial language learning paradigm in which the same previously unambiguous words were paired (across participants) with new related or unrelated meanings. One crucial advantage of this approach is that it allows researchers to “create” homonyms and polysemes in a well-controlled experimental setting, rather than class their word stimuli as such based on dictionary entries, subjective ratings, or linguistic criteria.

Further research into children's and adults' ability to learn new meanings for familiar words is imperative. Not only does such research provide a novel avenue for testing predictions and assumptions of the ambiguity literature, but it can also help us uncover the fundamental mechanisms underlying successful language learning. Although there has been some progress in understanding how children learn new words or new meanings for words they already know (e.g., Casenhiser, 2005; Doherty, 2004; Storkel & Maekawa, 2005), existing models of vocabulary acquisition have largely ignored learning multiple form-to-meaning mappings (see Dautriche, Chemla, & Christophe, 2016), and how we continually expand our vocabulary throughout the lifespan.

Chapter 6: Semantic Competition in Word Comprehension: Evidence from Polysemy

6.1 Introduction

The work discussed in Chapter 5 demonstrates that meaning relatedness is an important property of ambiguity that has a pervasive influence on both learning and processing word meanings. While these findings make a valuable contribution to the research on the relatedness effect, or, more specifically, the distinction between homonymy and irregular polysemy, they shed little light on other forms of polysemy. It remains unclear, for example, whether metaphor and metonymy would also slow word comprehension as a result of semantic competition, and whether these two differ from irregular polysemy. This was investigated in a further experiment involving existing ambiguous words.

Experiment 6 aimed to examine the processing of ambiguity, in particular polysemy, in natural language by contrasting relatedness decisions to irregular, metaphorical, and metonymic polysemes (for the properties and examples of the polysemes, see Table 6.1 below). Unambiguous words served as a control condition in order to determine whether any of the polysemes produce a processing disadvantage. Balanced homonyms served as an additional control condition, in that any disadvantage comparable to that for the homonyms would indicate separate semantic representations and maximum competition.

Table 6.1: Properties and examples of irregular, metaphorical, and metonymic polysemes.

Polyseme	Defining properties	Examples
Irregular	the alternative sense is loosely related (to the dominant sense), literal, and does not follow conventional rules of sense extension	"military/water tank" "restaurant/website menu"
Metaphorical	the alternative sense is loosely related, figurative, and does not follow conventional rules of sense extension	"wooden/authoritative chair" "swollen/table leg"
Metonymic	the alternative sense is highly related, literal, and follows conventional rules of sense extension (e.g., animal for meat alternations)	"fluffy/marinated rabbit" "dust/to dust"

Several steps were taken to ensure accurate manipulation of ambiguity type in the present experiment. Homonyms, irregular polysemes, metaphors, and metonyms were selected based on linguistic criteria. Having unrelated meanings for homonyms, loosely related senses for irregular polysemes and metaphors, and highly related senses for metonyms were then confirmed based on subjective ratings of relatedness between meanings/senses. All words were also controlled for the number of related senses in a dictionary. This experiment is, therefore, the first to combine the linguistic, subjective, and lexicographic approaches to the operalisation of ambiguity, with the potential to uncover whether the impact of polysemy in word processing is driven by the nature of sense extension, sense relatedness, or sense count. As discussed in the introduction to Experiment 5, there is disagreement in the literature on how polysemy might influence word comprehension in the absence of contextual bias, depending on how one defines and manipulates this form of ambiguity. Studies of subjective and linguistic polysemy predict a disadvantage effect for irregular polysemes and metaphors that have loosely related senses (e.g., Brown, 2008; Klepousniotou et al., 2008;

Lopukhina et al., 2018; cf. Brocher et al., 2016, 2018), but no such effect for metonyms that have highly related senses (e.g., Frazier & Rayner, 1990; Frisson, 2009; Klepousniotou, 2002). In contrast, studies of lexicographic polysemy predict a disadvantage effect for all polysemes, defined as many-sense words, regardless of how and to what extent their different senses are related (e.g., Beretta et al., 2005; Rodd et al., 2002). Therefore, Experiment 6 provides a way to test these contradictory predictions and inform future studies on how to best capture the essence of polysemy.

It is important to note that this investigation is also the first to compare all forms of polysemy (i.e., irregular, metaphorical, and metonymic) in a single experiment. Although there have been a few notable studies on differences within polysemy, these focussed on metaphors and metonyms and never included irregular polysemes (e.g., Jager & Cleland, 2015; Klepousniotou & Baum, 2007; MacGregor et al., 2012). Extending this line of research, the present experiment aimed to delineate how readers access and disambiguate these three types of words out of context. For instance, while some assume that irregular polysemy and metaphor belong to the same category (Apresjan, 1974; Brocher et al., 2016), research is yet to demonstrate that the two are indeed processed in a similar manner, and that the alternative sense being literal (in irregular polysemes) or figurative (in metaphors) makes no difference to the language system. Inclusion of irregular polysemes was also motivated by the need to validate the findings from Experiment 5. Evidence from learning new meanings suggests that irregular polysemes have separate but overlapping representations that compete with each other, whereas evidence from existing ambiguous words (Brocher et al., 2016, 2018) tends to support the single-representation account.

Thus, it was important to substantiate that competition for irregular polysemes was not caused by the artificial language paradigm, and that it also occurs in natural language.

In summary, Experiment 6 was primarily designed to establish whether all forms of polysemy produce semantic competition, as predicted by PDP models of ambiguity processing (e.g., Kawamoto, 1993; Rodd et al., 2004). This was necessary to advance the account of the ambiguity disadvantage within this thesis and validate assumptions of the models. A subsidiary aim was to bridge studies of lexicographic (e.g., Armstrong & Plaut, 2016; Rodd et al., 2002), subjective (e.g., Hino et al., 2006; Klepousniotou et al., 2008), and linguistic polysemy (e.g., Brocher et al., 2018; Klepousniotou & Baum, 2007) by combining their respective manipulations of this form of ambiguity. This was expected to clarify which of these underlies the impact of polysemy and unravel key inconsistencies in the findings on differences between polysemy and homonymy as well as within polysemy in the current literature.

6.2 Experiment 6: Short Prime Duration

6.2.1 Method

6.2.1.1 Participants

Thirty students and members of staff [24 females, aged 18-34 (*M* = 21.1, *SD* = 3.6)] from the University of Leeds participated in the experiment in exchange for four course credits or £3. All participants were right-handed monolingual native speakers of British English with no known history of language-/vision-related difficulties/disorders. The experiment received ethical approval from the School of Psychology, University of Leeds Ethics Committee.

6.2.1.2 Stimuli

The set of primes in this experiment comprised 28 balanced homonyms, 28 irregular polysemes, 28 metaphors, 28 metonyms, and 28 unambiguous control words. Sixteen of the homonyms were used in Experiments 2-4, whereas the other were taken from Klepousniotou and Baum (2007) and Klepousniotou et al. (2012)[30]. Balanced meaning frequencies for the latter group was confirmed based on the American-English eDom norms (Armstrong et al., 2012). Six of the homonyms had a third LF meaning (e.g., “tick” denoting the casing for a mattress) in the Wordsmyth Dictionary (Parks et al., 1998), while another two (“nail”, “organ”) had only one entry in the dictionary despite having two salient unrelated word meanings. Most of the homonyms had noun-noun (n = 15) or noun-verb interpretations (n = 9), while the rest had noun-adjective (n = 3) or adjective-adjective interpretations (n = 1).

A number of linguistic criteria were used to compile the list of irregular polysemes. First, unlike homonyms, the senses of these words were related through a single semantic feature, such as physical property (“military/water tank”) or function (“restaurant/website menu”). Second, though semantically related, the multiple senses of irregular polysemes could not be derived via a

---

[30] A new set of balanced homonyms was necessary to match these words to other primes for important word properties (e.g., word frequency, length, imageability).

productive rule typical of metonymy, such as animal-meat (e.g., “duck”) or object-action alternations (e.g., “shield”). Finally, unlike metaphors, both the dominant and the alternative/subordinate sense of these words were strictly literal. Most of the selected irregular polysemes had noun-noun (*n* = 16) or noun-verb interpretations (*n* = 7), with a small number of words with verb-verb (*n* = 3), adjective-adjective (*n* = 2), and noun-adjective interpretations (*n* = 1).

Metaphors had loosely related senses that could not be derived via a productive rule typical of metonymy either. Unlike irregular polysemes, however, the subordinate senses of metaphors were figurative rather than literal (e.g., “lion” as the animal or a brave person) and exhibited a number of predictable relations – namely, animal for human trait (e.g., “lion”; *n* = 12), object for human trait (e.g., “pearl”; *n* = 5), object for action (e.g., “grill”; *n* = 5), and body part for object alternations (e.g., “leg”; *n* = 3). There were only three metaphors with truly idiosyncratic patterns of sense extension (e.g., “cult” denoting a religious group or a craze). As in the other sets of primes, metaphors had noun-noun (*n* = 22) or noun-verb interpretations (*n* = 4), with only two words with noun-adjective interpretations (*n* = 2).

All metonyms had highly related, literal interpretations that followed common patterns of productive sense extension (for a list of these patterns, see Srinivasan & Rabagliati, 2015). Like metaphors, most words were taken from Klepousniotou and Baum (2007) and Klepousniotou et al. (2012). Metonymic polysemes exhibited five different types of relations – producer for product (e.g., “rabbit”; *n* = 11), object for colour (e.g., “silver”; *n* = 5), count for mass (e.g., “carrot”; *n* = 5), substance for action involving taking that substance (e.g., “skin”;

n = 4), and container for contents alternations (e.g., "tin"; *n* = 3). The words had noun-noun (*n* = 15), noun-adjective (*n* = 9), or noun-verb interpretations (*n* = 4).

Unlike homonyms, each irregular polyseme, metaphor, and metonym had a single meaning/entry in the Wordsmyth Dictionary (Parks et al., 1998). The three sets of polysemous prime words were matched for the number of word senses in the dictionary (see stimulus properties in Table 6.2 below; all *p*s = 1[31]). Since homonyms had two unrelated meanings, sense count for these words was almost twice as high as that for polysemous words (all *p*s ≤ .01). The four types of ambiguous primes were also statistically comparable in terms of semantic diversity (all *p*s > .05). Finally, the stimulus set comprised 28 unambiguous nouns with either one (*n* = 26) or two senses (*n* = 2) in the Wordsmyth Dictionary (Parks et al., 1998). The number of senses was naturally lower for unambiguous than ambiguous primes (all *ps* < .001). Likewise, the ratings of semantic diversity for unambiguous words were lower than those for balanced homonyms, irregular polysemes, and metaphors (all *p*s < .05), but not metonyms (*p* = 1). The five sets of primes were matched (all *F*s ≤ 2) at the group level with respect to other word properties such as word-form frequency, imageability, and orthographic neighbourhood size (see Table 6.2 below). None of the prime words was a homophone or a compound word.

Thirty monolingual native speakers of British English [12 females; aged 18-40 (*M* = 27.9, *SD* = 6.3)] rated meaning relatedness in an online rating study. This pre-test was conducted for all ambiguous words and aimed to confirm that they involved different degrees of meaning/sense relatedness. Participants read

---

[31] Throughout this subsection, *p*-values refer to pairwise comparisons (with the Bonferroni adjustment) that were conducted for significant one-way between-items ANOVAs with the factor of Prime/Target.

definitions of the dominant and subordinate senses of the words and rated their semantic relatedness on a 7-point scale (where 1 denoted "highly unrelated" and 7 denoted "highly related"). The definitions were taken from the Wordsmyth Dictionary (Parks et al., 1998) and were presented in pseudo-randomised order. The manipulation of sense relatedness was successful [see Figure 6.2 below; $F(3, 57.5) = 398.2$, $p < .001$]; the ratings were significantly lower for homonyms than for any other primes (all *p*s < .001). For polysemes, sense relatedness was greater for metonyms than irregular polysemes ($p < .001$) and metaphors ($p < .001$), with no significant difference between the latter two ($p = 1$). All primes and their mean relatedness ratings are given in Appendix 5.

Table 6.2: Experiment 6: Properties of the prime words.

Variable	Balanced homonyms	Irregular polysemes	Metaphors	Metonyms	Unambiguous words
Example	“fan”	“spoil”	“lion”	“gold”	“lake”
Meaning/sense relatedness	2.0 (0.4)	4.5 (1.1)	4.6 (0.9)	6.1 (0.5)	-
Senses	9.4 (5.3)	5.4 (2.1)	5.1 (2.3)	5.1 (2.7)	1.1 (0.3)
Semantic diversity	1.6 (0.2)	1.6 (0.2)	1.6 (0.1)	1.5 (0.2)	1.4 (0.2)
Letters	4.6 (1.1)	5.0 (0.8)	4.6 (1.1)	5.0 (1.2)	4.8 (1.1)
Phonemes	3.8 (1.0)	4.0 (0.9)	3.8 (0.9)	4.1 (1.2)	4.0 (1.1)
Syllables	1.3 (0.4)	1.4 (0.5)	1.3 (0.5)	1.5 (0.6)	1.6 (0.5)
Raw frequency	21.6 (16.5)	21.7 (14.0)	21.8 (23.7)	21.7 (23.3)	21.1 (13.9)
Log frequency	1.2 (0.3)	1.2 (0.3)	1.1 (0.4)	1.1 (0.5)	1.2 (0.3)
Orthographic neighbours	7.2 (6.2)	4.0 (3.9)	4.7 (4.0)	4.0 (5.1)	4.1 (4.1)
Log bigram frequency	2.8 (0.4)	2.7 (0.4)	2.7 (0.4)	2.7 (0.4)	2.6 (0.6)
Imageability	5.5 (0.7)	5.6 (0.6)	5.7 (0.9)	5.9 (0.4)	5.7 (0.6)
Concreteness	5.7 (0.3)	5.7 (0.3)	5.8 (0.8)	6.0 (0.3)	5.7 (0.7)
Subjective familiarity	5.2 (0.5)	5.2 (0.6)	5.2 (0.6)	5.4 (0.5)	5.2 (0.6)

Note: Standard deviations are given in the parentheses. Each prime set contained 28 items. “Senses” refer to the number of related word senses in the Wordsmyth Dictionary (Parks et al., 1998). Semantic diversity values come from Hoffman et al. (2013). Word-form frequency, bigram frequency, and the number of orthographic neighbours come from the BNC (2007). Imageability, concreteness, and familiarity ratings come from the MRC Database (Coltheart, 1981).

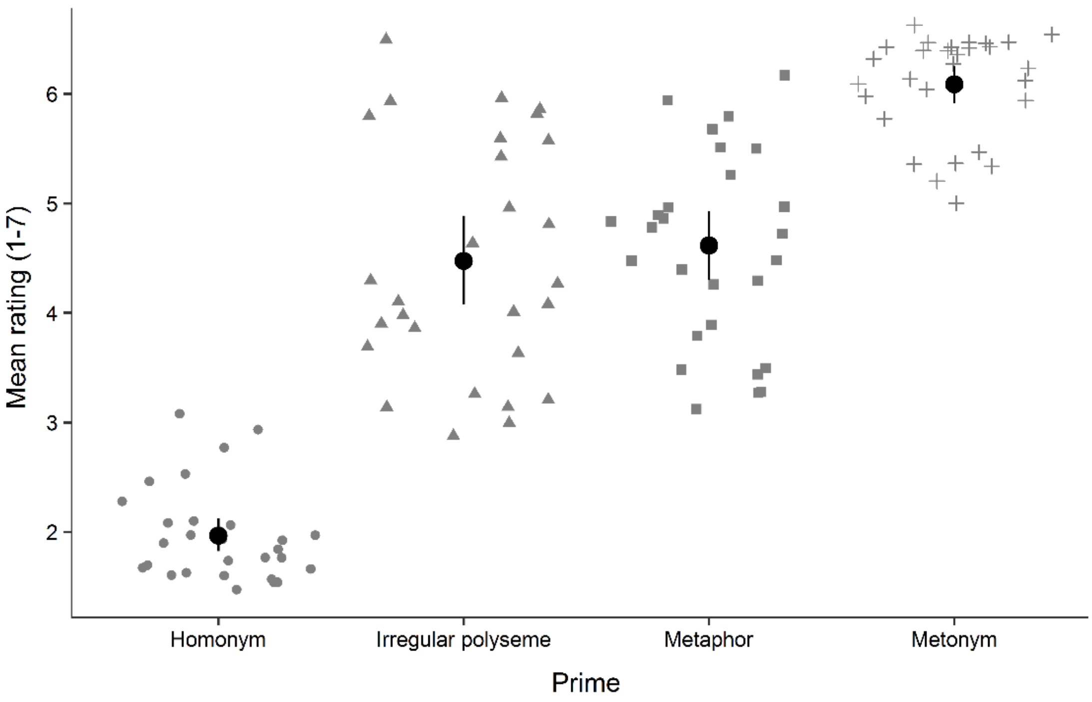

Figure 6.1: Experiment 6: Scatterplot of item mean meaning-/sense-relatedness ratings across the four sets of ambiguous primes. Relatedness was rated on a 7-point scale where 1 denoted "highly unrelated" and 7 denoted "highly related". The black circles and error bars show the mean rating and 95% CIs for each set, respectively.

Each prime was paired with two targets that related either to the dominant (e.g., "violet-tulip") or the subordinate sense (e.g., "violet-purple") and one unrelated target (e.g., "violet-roar"), such that all participants responded to each prime word three times. For balanced homonyms, the "dominant" meaning was the one with the slightly higher frequency rating in the British (Experiment 1) or American eDom norms (Armstrong et al., 2012). Since there are not such norms for polysemous words, sense dominance for irregular polysemes, metaphors, and metonyms was determined based on entry order in the Wordsmyth Dictionary (Parks et al., 1998) - the dominant sense was the one listed first. The dictionary lists related senses in the order of frequency of use, which has been shown to closely correspond to native speakers' ratings of sense dominance (see Jager & Cleland, 2016). For unambiguous words, the dominant- and the subordinate-sense target referred to the same interpretation of the word (e.g., “pond-puddle” vs. “pond-swamp”). Targets related to primes through category membership (e.g., “gold-pink”), action-recipient relationship (e.g., “orange-squeeze”), physical properties (e.g., “hood-hat”), and function (e.g., “butler-servant”). None of the related targets instantiated both senses of ambiguous words.

All targets were heterophonic non-homonyms, and their low numbers of word senses in the Wordsmyth Dictionary (Parks et al., 1998) were matched [$F(14, 405) = 0.1$, $p = 1$] across the 15 conditions (see stimulus properties in Tables 6.3-6.5 below). The syntactic class of the targets was controlled as much as possible, such that each set contained 15-26 nouns, 2-6 verbs, and 0-7 adjectives. All but one target (“website”) were non-compound words. Prime-target word pairs were also statistically comparable (all Fs ≤ 1.3) with respect to the lexical and semantic variables used to match the prime words (see Tables 6.3-6.5 below), except for concreteness [$F(14, 90.1) = 2.7$, $p < .01$], imageability [$F(14, 291) = 2.0$, $p < .05$], and semantic diversity [$F(14, 399) = 1.8$, $p < .05$]. Concreteness ratings for the subordinate-sense targets of metaphors (e.g., “lion-brave”) were lower than those for every other, otherwise well-matched, set (all ps < .05). That set also had lower imageability ratings, but only relative to the dominant- and subordinate-sense targets of unambiguous words and all targets of metonyms (all ps < .05). In contrast, semantic diversity ratings for the subordinate-sense targets of metaphors were significantly higher than those for the dominant- and subordinate-sense targets of metonyms (both ps < .05).

A different group of 30 monolingual native speakers of British English [15 females; aged 18-40 ($M = 28.8$, $SD = 6.8$)] rated the semantic relatedness between the prime and the target using a 7-point scale (where 1 denoted “highly unrelated” and 7 denoted “highly related”). The word stimuli were divided into 10 pseudo-randomised blocks containing 28 related and 14 unrelated pairs. The order of blocks was counter-balanced across raters. The pre-test confirmed that the prime-target relatedness ratings for unrelated pairs [$F(4, 135) = 1.2$, $p = .32$] did not reliably differ across the five sets of prime words (see Table 6.5 below). For related pairs, the ratings for the subordinate-sense targets of irregular polysemes and metaphors were significantly lower than those for most other, otherwise well-matched, sets (see Tables 6.3 & 6.4 below).

A different group of 30 monolingual British-English native speakers [16 females, aged 20-39 ($M = 31.0$, $SD = 5.3$)] participated in a follow-up prime-target relatedness pre-test. The aim of this study was to verify whether low relatedness ratings for some of the polysemes and their subordinate-sense targets (e.g., “summit-diplomat”) truly reflected poor conceptual overlap, or whether participants’ ratings were biased by sense dominance. In this online pre-test, participants read a sentence (all taken from the Oxford Dictionary, Simpson & Weiner, 1989) containing a polyseme in its subordinate sense (e.g., “These days, representatives of China, Japan and South Korea attend all summits and ministerial meetings.”) and then used the same 7-point scale to rate the relatedness between the polysemous prime and four words – namely, two unrelated filler words and two targets related to the subordinate sense. Critically, these relatedness ratings were collected for all irregular, metaphorical, and metonymic polysemes.

Table 6.3: Experiment 6: Properties of the dominant-sense targets across the five sets of primes.

Variable	Balanced homonyms	Irregular polysemes	Metaphors	Metonyms	Unambiguous words
Example	“fan-cheer”	“spoil-ruin”	“lion-jaguar”	“gold-bronze”	“pond-puddle”
Prime-target relatedness	6.0 (0.4)	6.0 (0.6)	6.2 (0.5)	6.0 (0.5)	6.2 (0.5)
Senses	4.6 (2.8)	4.5 (2.3)	4.6 (2.7)	4.6 (2.7)	4.3 (2.7)
Semantic diversity	1.6 (0.3)	1.6 (0.2)	1.5 (0.2)	1.5 (0.3)	1.5 (0.2)
Letters	5.0 (1.2)	5.3 (1.3)	5.0 (1.4)	5.3 (1.2)	5.2 (1.1)
Phonemes	4.0 (1.1)	4.5 (1.1)	4.4 (1.4)	4.5 (1.1)	4.1 (1.3)
Syllables	1.4 (0.6)	1.7 (0.7)	1.7 (0.7)	1.5 (0.6)	1.5 (0.6)
Raw frequency	19.2 (17.9)	19.9 (17.7)	19.0 (23.8)	19.9 (26.7)	19.5 (18.4)
Log frequency	1.1 (0.5)	1.1 (0.4)	1.0 (0.5)	1.0 (0.6)	1.1 (0.5)
Orthographic neighbours	4.5 (4.4)	3.9 (5.5)	4.0 (5.7)	4.0 (4.1)	4.5 (6.0)
Log bigram frequency	2.8 (0.4)	2.5 (0.4)	2.6 (0.4)	2.6 (0.6)	2.8 (0.4)
Imageability	5.3 (1.0)	5.3 (1.1)	5.4 (1.0)	5.4 (1.0)	5.7 (0.9)
Concreteness	5.5 (1.0)	5.6 (0.9)	5.3 (1.0)	5.8 (0.6)	5.7 (0.4)
Subjective familiarity	5.3 (0.5)	5.3 (0.5)	5.2 (0.6)	5.3 (0.7)	5.1 (0.7)

Note: Standard deviations are given in the parentheses. Each set contained 28 items. Information on the different variables is given in the note for Table 6.2.

Table 6.4: Experiment 6: Properties of the subordinate-sense related targets across the five sets of primes.

Variable	Balanced homonyms	Irregular polysemes	Metaphors	Metonyms	Unambiguous words
Example	“fan-rotate”	“spoil-brat”	“lion-brave”	“gold-pink”	“pond-swamp”
Prime-target relatedness	5.9 (0.5)	5.4 (1.1)	5.0 (1.1)	5.6 (1.0)	6.2 (0.5)
Senses	4.5 (3.1)	4.3 (2.5)	4.6 (2.8)	4.5 (3.0)	4.6 (3.1)
Semantic diversity	1.5 (0.3)	1.6 (0.2)	1.7 (0.2)	1.5 (0.3)	1.5 (0.3)
Letters	5.1 (1.2)	5.4 (1.4)	5.4 (1.7)	5.0 (1.2)	5.0 (1.2)
Phonemes	4.2 (1.3)	4.6 (1.5)	4.7 (1.6)	4.2 (1.4)	4.3 (1.2)
Syllables	1.6 (0.6)	1.7 (0.7)	1.7 (0.7)	1.5 (0.5)	1.5 (0.5)
Raw frequency	19.9 (22.9)	20.9 (31.2)	20.7 (21.5)	18.9 (29.1)	19.8 (24.0)
Log frequency	1.0 (0.5)	1.0 (0.6)	1.1 (0.6)	1.0 (0.5)	1.0 (0.5)
Orthographic neighbours	4.4 (5.5)	3.9 (4.3)	4.4 (4.9)	4.6 (5.2)	4.5 (4.6)
Log bigram frequency	2.7 (0.4)	2.7 (0.4)	2.6 (0.6)	2.7 (0.4)	2.6 (0.6)
Imageability	5.4 (0.6)	5.5 (0.7)	4.4 (1.0)	5.6 (0.6)	5.6 (0.8)
Concreteness	5.4 (0.7)	5.2 (0.8)	3.9 (1.3)	5.4 (0.7)	5.4 (1.0)
Subjective familiarity	5.2 (0.5)	5.6 (0.5)	5.2 (0.4)	5.3 (0.6)	5.1 (0.7)

Note: Standard deviations are given in the parentheses. Each set contained 28 items. Information on the different variables is given in the note for Table 6.2.

Table 6.5: Experiment 6: Properties of the unrelated targets across the five sets of primes.

Variable	Balanced homonyms	Irregular polysemes	Metaphors	Metonyms	Unambiguous words
Example	“fan-jaw”	“spoil-jug”	“lion-pirate”	“gold-penguin”	“pond-wipe”
Prime-target relatedness	1.7 (0.3)	1.8 (0.3)	1.7 (0.2)	1.7 (0.3)	1.7 (0.3)
Senses	4.5 (2.6)	4.6 (2.8)	4.6 (2.5)	4.3 (2.5)	4.3 (2.8)
Semantic diversity	1.5 (0.3)	1.5 (0.2)	1.5 (0.4)	1.5 (0.2)	1.5 (0.3)
Letters	5.0 (1.3)	5.1 (1.4)	5.3 (1.2)	5.3 (1.1)	5.0 (1.3)
Phonemes	4.4 (1.3)	4.1 (1.3)	4.6 (1.3)	4.4 (1.3)	4.2 (1.5)
Syllables	1.6 (0.6)	1.4 (0.6)	1.7 (0.7)	1.5 (0.6)	1.5 (0.7)
Raw frequency	19.5 (22.9)	20.2 (21.5)	19.4 (22.3)	20.6 (20.1)	20.2 (29.3)
Log frequency	1.0 (0.5)	1.0 (0.6)	1.0 (0.5)	1.1 (0.5)	0.9 (0.7)
Orthographic neighbours	4.3 (5.3)	4.4 (5.3)	4.1 (5.2)	4.1 (3.9)	4.3 (4.1)
Log bigram frequency	2.5 (0.4)	2.5 (0.7)	2.7 (0.4)	2.7 (0.4)	2.8 (0.6)
Imageability	5.3 (0.9)	5.2 (0.9)	5.5 (1.0)	5.6 (0.7)	5.3 (0.9)
Concreteness	5.5 (0.8)	5.2 (1.1)	5.6 (0.7)	5.4 (1.1)	5.3 (1.1)
Subjective familiarity	5.1 (0.4)	5.2 (0.6)	5.1 (0.6)	5.2 (0.8)	5.2 (0.6)

Note: Standard deviations are given in the parentheses. Each set contained 28 items. Information on the different variables is given in the note for Table 6.2.

The “primed” ratings (all at or above 4.9) for the subordinate-sense targets of all polysemous words were significantly higher (irregular polysemes: $M = 6.1$, $SD = 0.4$; metaphors: $M = 6.0$, $SD = 0.5$; metonyms: $M = 5.9$, $SD = 0.5$) than those in the previous “unprimed” pre-test [irregular polysemes: $M = 5.4$, $SD = 1.1$; metaphors: $M = 5.0$, $SD = 1.1$; metonyms: $M = 5.6$, $SD = 1.0$; $t(83) = 8.0$, $p < .001$] and no longer significantly lower than the ratings for the other sets of related pairs [$F(9, 270) = 1.0$, $p = .48$]. These results indicate that the subordinate-sense targets of polysemes were indeed related and suitable for experimental testing. As with unbalanced homonyms in the LF meaning (see 3.2.1.2 Stimuli in Chapter 3), the subordinate-sense targets of polysemes were given relatively lower relatedness ratings (in the original stimulus pre-test) because the raters did not sufficiently retrieve the less salient sense when these words were presented in isolation.

Finally, the stimulus set also included 140 unrelated fillers that did not contain any of the experimental primes or targets. These word pairs served to balance the number of related and unrelated responses. All experimental word pairs and their mean relatedness ratings are given in Appendix 6.

6.2.1.3 Procedure

The relatedness decision task and response output were the same as in Experiments 2, 3, and 5 (for the rationale behind the task, see 1.5 Aims of the Thesis in Chapter 1). The 15 sets of word pairs ($N = 420$) and 140 unrelated fillers were presented in four pseudo-randomised blocks. Each block contained the same number of related and unrelated trials, the same number of the five types of prime words, and the same number of the three types of experimental target words (dominant-sense, subordinate-sense, unrelated). None of the primes appeared more than once within the same block. The order of trials in each block was pseudo-randomised, such that no more than three related/unrelated pairs appeared consecutively. The task began with a 20-trial practice block comprising different examples of conditions. In the practice phase, participants were given feedback on both the speed and accuracy of their responses. 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 items (not included in the analysis) that aimed to help them get back to the habit of quick response-making.

Both the trial sequence and stimulus duration were the same as in Experiments 2 and 5. That is, trials began with a 500 ms fixation cross. After a delay of 100 ms, the prime and the target were presented for 200 ms and 500 ms, respectively, with a delay of 50 ms in between. Once the target word disappeared, there was a 1500 ms delay for response execution followed by a 100 ms ITI. Participants could make relatedness decisions as soon as the target word was presented, 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). Response speed and accuracy were equally emphasised in the instructions, and participants were told what constituted semantic relatedness and given examples. On average, testing lasted for 28 minutes.

6.2.2 Results

Three of the 30 participants were removed from all analyses – two due to a large number of responses that were not made within 1500 ms (27.3/21.5% of all responses) and one due to very slow and variable performance across all trials (*M* = 854.7 ms, *SD* = 293.7). For RTs, analyses excluded errors (21.9% of all trials) and outliers (two standard deviations above/below a participant's mean per condition; 4.0% of the remaining trials). RTs were log-transformed to normalise the distribution of residuals. Back-transformed means of RTs and CIs that were estimated from mixed-effects models are reported below.

Accuracy and latency data were analysed using logit or linear mixed-effects models. The models included the factors of Prime (homonym, irregular polyseme, metaphor, metonym, unambiguous word) and Target (dominant-sense, subordinate-sense, unrelated). As in Experiments 2 and 3, response-latency models also included the factor of Block (1, 2, 3, 4). Nevertheless, main effects and interactions involving Block are not reported as the sole purpose of this factor was to account for potential practice or prime-repetition effects (Pollatsek & Well, 1995), and no such effects were detected in either of the experiments reported below[32]. Terms involving Block were removed from accuracy analyses because they resulted in model non-convergence. Each model included random intercepts for subjects and items. The random slope for the effect of Block across subjects was significant and included in the response-latency but not the response-accuracy model. All planned contrasts were

[32] Consistent with Experiments 2-5, analyses for Experiments 6 and 7 demonstrated that participants' performance did not reliably change across the blocks of these short experiments. Likewise, there was no indication that having responded to an ambiguous word in one meaning affected the processing of that word on subsequent trials instantiating the other meaning.

adjusted using the Holm-Bonferroni method. Contrasts for the Prime × Target interaction compared the effects of Prime for each target type.

6.2.2.1 Response Accuracy

Mean error rates across the conditions are illustrated in Figure 6.2 below. The response-accuracy model (marginal R^2 = .20, conditional R^2 = .49) revealed a significant main effect of Prime [$\chi^2(4) = 51.4$, $p < .001$]. Relative to unambiguous words (M = 10.8%, SD = 5.5), error rates were higher for homonyms (M = 28.7%, SD = 8.8; $p < .001$), irregular polysemes (M = 25.1%, SD = 7.8; $p < .001$), and metaphors (M = 27.8%, SD = 7.8; $p < .001$), but not metonyms (M = 17.0%, SD = 7.3; $p = .20$). Error rates for homonyms ($p < .001$), irregular polysemes ($p < .05$), and metaphors ($p < .01$) were also higher than those for metonyms. Further, there was a significant main effect of Target [$\chi^2(2) = 121.6$, $p < .001$]. Responses were more erroneous to subordinate-sense (M = 36.2%, SD = 10.0) than dominant-sense (M = 21.4%, SD = 7.9; $p < .001$) and unrelated targets (M = 8.0%, SD = 4.5; $p < .001$), with a significant difference between the latter two ($p < .001$).

The response-accuracy model also revealed a significant interaction between the effects of Prime and Target [$\chi^2(8) = 48.8$, $p < .001$]. For dominant-sense targets, planned contrasts indicated higher error rates for homonyms (M = 33.5%, SD = 10.3; $p < .001$), irregular polysemes (M = 29.1%, SD = 11.5; $p < .001$), and metonyms (M = 20.0%, SD = 9.5; $p < .05$) than unambiguous words (M = 8.9%, SD = 7.0). Responses were also more erroneous to homonyms than metaphors (M = 15.7%, SD = 9.2; $p < .01$). In contrast, for subordinate-sense targets, planned contrasts indicated higher error rates for metaphors (M = 60.3%, *SD* = 13.7) than homonyms (*M* = 41.9%, *SD* = 15.2; *p* < .05), irregular polysemes (*M* = 39.4%, *SD* = 11.9; *p* < .01), metonyms (*M* = 24.7%, *SD* = 11.6; *p* < .001), and unambiguous words (*M* = 14.7%, *SD* = 8.9; *p* < .001). Relatedness decisions were also more erroneous to homonyms compared to metonyms (*p* < .01) and unambiguous words (*p* < .001), as well as to irregular polysemes compared to metonyms (*p* < .01) and unambiguous words (*p* < .001). For unrelated targets, there were no significant differences among the types of prime words (all *ps* > .32).

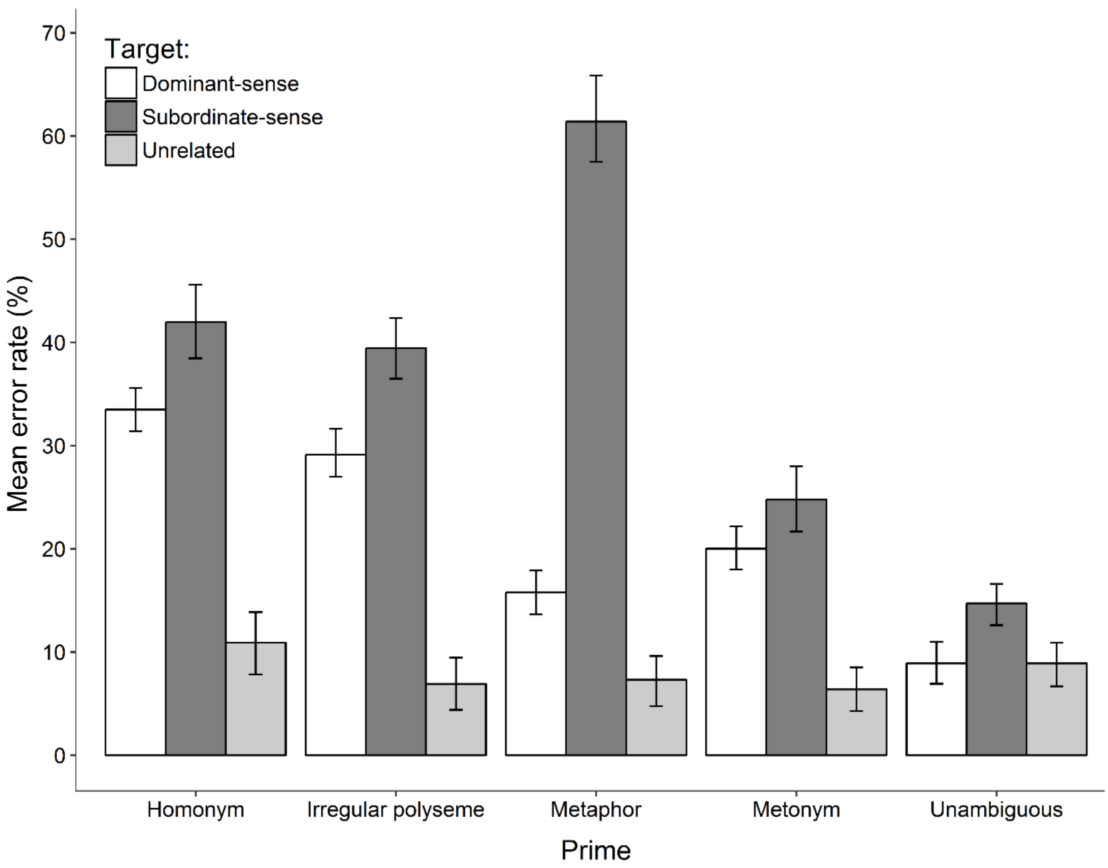

Figure 6.2: Experiment 6: Subject means of error rates across the conditions of Prime and Target. Error bars show 95% CIs adjusted to remove between-subjects variance.

6.2.2.2 Response Latency

Mean RTs across the conditions are illustrated in Figure 6.5 below. The response-latency model (marginal R^2 = .07, conditional R^2 = .48) revealed a significant main effect of Prime [$\chi^2(4) = 81.2$, $p < .001$]. Relative to unambiguous words (*M* = 596.2 ms, 95% CIs: 564.3, 6330.1), RTs were higher for homonyms (*M* = 667.6 ms, 95% CIs: 631.7, 705.7; $p < .001$), irregular polysemes (*M* = 656.4 ms, 95% CIs: 621.2, 693.9; $p < .001$), metaphors (*M* = 641.8 ms, 95% CIs: 607.2, 678.4; $p < .001$), and metonyms (*M* = 620.2 ms, 95% CIs: 586.9, 655.4; $p < .01$). RTs were also higher for homonyms compared to metaphors ($p < .01$) and metonyms ($p < .001$), as well as irregular polysemes ($p < .001$) and metaphors ($p < .05$) compared to metonyms. There was a significant effect of Target [$\chi^2(2) = 56.2$, $p < .001$]. RTs were higher for unrelated (*M* = 657.8 ms, 95% CIs: 623.2, 694.4) than dominant-sense (*M* = 607.6 ms, 95% CIs: 575.6, 641.5; $p < .001$) and subordinate-sense targets (*M* = 642.4 ms, 95% CIs: 609.4, 679.4; $p < .05$), with a significant difference between the latter two ($p < .001$).

The latency model also showed a significant interaction between the effects of Prime and Target [$\chi^2(8) = 43.3$, $p < .001$]. For dominant-sense targets, planned contrasts indicated higher RTs for homonyms (*M* = 647.6 ms, 95% CIs: 609.3, 688.3) than metaphors (*M* = 593.2 ms, 95% CIs: 558.6, 629.9; $p < .05$), metonyms (*M* = 591.6 ms, 95% CIs: 556.9, 628.3; $p < .05$), and unambiguous words (*M* = 572,1 ms, 95% CIs: 538.9, 607.6; $p < .001$). Responses were also slower to irregular polysemes (*M* = 636.9 ms, 95% CIs: 599.5, 676.9) than metaphors ($p < .01$), metonyms ($p < .01$), and unambiguous control words ($p < .001$). In contrast, for subordinate-sense targets, planned contrasts indicated higher RTs for homonyms (*M* = 668.3 ms, 95% CIs: 628.6, 710.6; $p < .001$), irregular polysemes (*M* = 660.8 ms, 95% CIs: 621.4, 702.9; $p < .001$), metaphors

(*M* = 696.5 ms, 95% CIs: 653.9, 741.7; $p < .001$), and metonyms (*M* = 623.0 ms, 95% CIs: 586.5, 661.9; $p < .01$) compared to unambiguous words (*M* = 575.3 ms, 95% CIs: 541.8, 611.1). Responses were also slower to homonyms ($p < .01$), irregular polysemes ($p < .05$), and metaphors ($p < .001$) compared to metonyms. For unrelated targets, planned contrasts indicated higher RTs for homonyms (*M* = 687.5 ms, 95% CIs: 647.4, 730.0) than metaphors (*M* = 639.9 ms, 95% CIs: 602.6, 679.4; $p < .01$), metonyms (*M* = 647.3 ms, 95% CIs: 609.7, 687.2; $p < .05$), and unambiguous words (*M* = 643.9 ms, 95% CIs: 606.5, 683.8; $p < .05$).

As in Experiments 2 and 3, further analysis of RTs included properties of the prime and the target as covariates. The rationale was that due to a large number of errors for metaphors in the subordinate sense, this condition in the above analysis involved only a subset of items that was no longer well-matched with its counterparts. In fact, the subordinate-sense targets of metaphors differed from other sets of word pairs in terms of imageability, concreteness, and semantic diversity ratings even at the stimulus-matching stage (see 6.2.1.2 Stimuli above). Each of the variables used for the matching of primes (see Table 6.2 above) and targets (see Tables 6.3-6.5 above) was considered for inclusion, except for those related to the manipulations within Prime and Target (i.e., sense relatedness, number of senses, and semantic diversity at the prime level and prime-target relatedness at the target level). In order to prevent model over-fitting, analysis included only those variables that significantly correlated with mean RT - syntactic ambiguity (i.e., whether the different interpretations of words corresponded to the same parts of speech) and imageability at the prime level as well as imageability at the target level[33]. All of the effects reported above were still significant when these variables were taken into account, except for slower responses to irregular polysemes than metaphors on dominant-sense trials, slower responses to irregular polysemes than metonyms on dominant- and subordinate-sense trials, and slower responses to homonyms than metonyms on unrelated trials (see Figure 6.3 below) that were no longer significant. Overall, then, this analysis shows that the results were indeed due to the manipulated properties of ambiguous words.

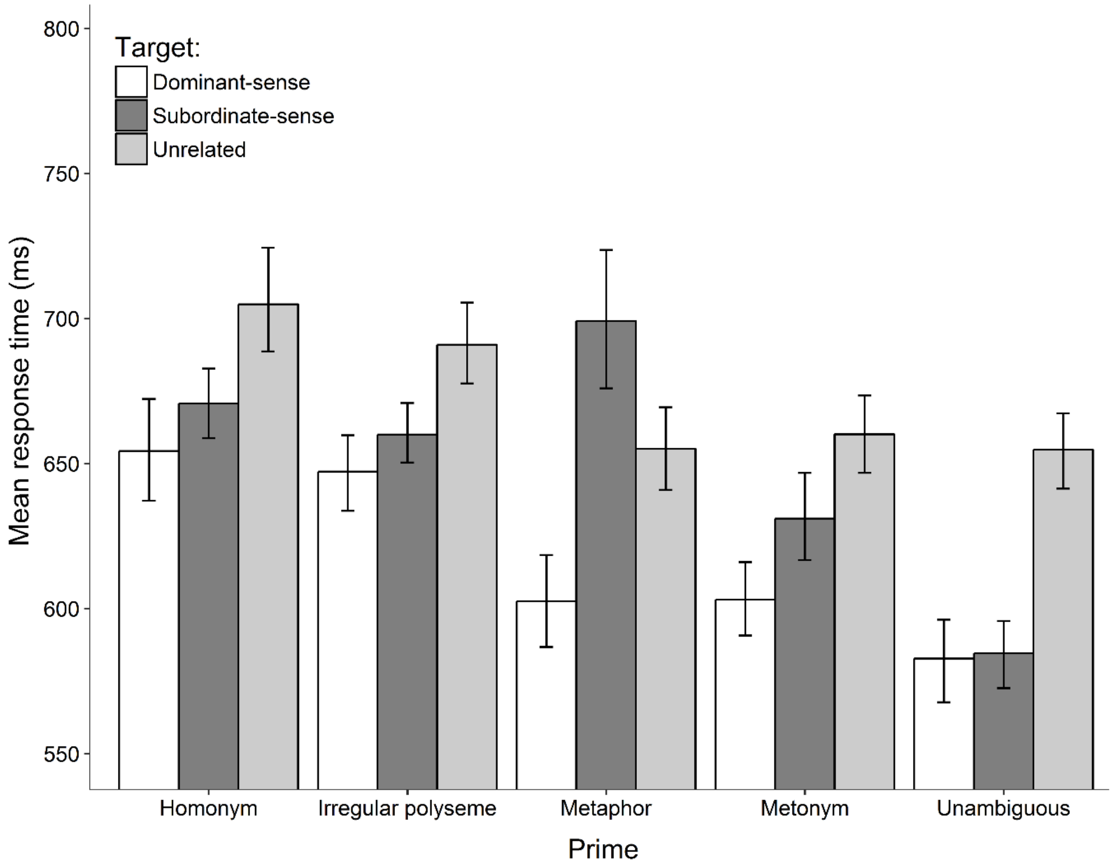

Figure 6.3: Experiment 6: Subject means of untransformed RTs across the conditions of Prime and Target. Error bars show 95% CIs adjusted to remove between-subjects variance.

---

[33] As in Experiments 2 and 3, imageability was chosen over concreteness due to the greater availability of item ratings in the MRC Psycholinguistic Database (Coltheart, 1981).

6.2.3 Discussion

Experiment 6 revealed marked differences in the processing of irregular, metaphorical, and metonymic polysemes. The overall ambiguity disadvantage (i.e., slower relatedness decisions to ambiguous than unambiguous words across all types of trials/targets) appeared only for balanced homonyms and irregular polysemes, though for the latter the slowing on unrelated trials (36.2 ms, relative to unambiguous control words) was significant before ($p < .05$) but not after the Holm-Bonferroni correction ($p = .28$). These results replicate the disadvantage effect for balanced homonyms (Experiments 2 & 3) in a new sample of words as well as the disadvantage effect for irregular polysemes (Experiment 5) in existing words. Interestingly, while Experiment 5 showed that newly-learnt meanings of irregular polysemes produced less competition than newly-learnt meanings of homonyms, Experiment 6 found no such effect in natural language. Nevertheless, the finding of competition for irregular polysemes, reduced or not, presents the shared features account (Brocher et al., 2016, 2018) with a daunting task of explaining why any competition would arise if these words supposedly have only one representation.

The pattern of responses was different for metaphors and metonyms. These words incurred a processing cost only when disambiguated toward the subordinate sense. Together with the null effect on unrelated trials, the results demonstrate that neither metaphors nor metonyms show the disadvantage effect, most likely because they have a single semantic representation (e.g., Frisson, 2009; Klepousniotou et al., 2012). It appears, then, that the two forms of polysemy hinder comprehension only in relatively rare situations when the dominant sense turns out to be incorrect, forcing readers to engage in an additional, time-consuming process of retrieving or deriving the subordinate sense. The finding of a large processing cost on subordinate-sense trials for metaphorical polysemes (121.2 ms, relative to unambiguous control words) and a much weaker one for metonymic polysemes (47.7 ms) suggests that the two may differ in the nature and difficulty of execution of the process.

6.3 Experiment 7: Long Prime Duration

The main aim of Experiment 7 was to investigate the impact of prime duration on the comprehension of polysemes. In particular, it is possible that metaphors and metonyms in Experiment 6 were generally processed as fast as their unambiguous counterparts because the task did not provide enough time to retrieve the subordinate sense, thus preventing potential competition. To explore this possibility, Experiment 7 involved the same task and stimuli as Experiment 6, but the primes were presented for a longer period (700 ms).

The manipulation of prime duration was not expected to affect the pattern of responses to metonyms. Research has shown that the different senses of metonyms, when encountered in neutral or out of context, are initially activated to the same extent (Klepousniotou et al., 2012; see also Bott, Rees, & Frisson, 2016), and remain so long after the presentation of the word (Klepousniotou & Baum, 2005a; MacGregor et al., 2015). This suggests that the longer prime duration would neither facilitate nor impede access to the subordinate sense of metonyms. As for metaphors, the evidence is less clear-cut. ERP studies reported weaker activation levels for the subordinate than the dominant sense at the ISI of 50 ms (Klepousniotou et al., 2012), but equivalent levels at the ISI of 750 ms (MacGregor et al., 2012). Behavioural studies, on the other hand, found minimal activation of the subordinate sense at either the ISI of 100 ms or 1000 ms (Klepousniotou & Baum, 2005a). Since the time-course of access to the literal and figurative senses of metaphors has clearly not been established, Experiment 7 was necessary to confirm that these words do not show a disadvantage effect even when participants have more time to retrieve the subordinate sense (assuming that they do so in minimal context to begin with).

6.3.1 Method

6.3.1.1 Participants

A different group of 39 students and members of staff [37 females, aged 18-35 (*M* = 20.0, *SD* = 3.6)] from the University of Leeds participated in the experiment in exchange for four course credits or £3. All participants were right-handed monolingual native speakers of British English with no known history of language- or vision-related difficulties or disorders. The experiment received ethical approval from the School of Psychology, University of Leeds Ethics Committee.

6.3.1.2 Stimuli & Procedure

The same stimuli and procedure as in Experiment 6 were used. The only difference between the two experiments was the longer prime-word duration in

Experiment 7 (700 ms instead of 200 ms). On average, testing lasted for 30 minutes.

6.3.2 Results

Four of the 39 participants were removed from all analyses – two due to a relatively large number of errors (39.0% of all trials in both cases) and the other two due to very slow performance across all trials (*M* = 930.5 ms, *SD* = 225.1; *M* = 971.7 ms, *SD* = 232.2). For RTs, analyses excluded errors (22.0% of trials) and outliers (two standard deviations above/below a participant's mean per condition; 4.5% of the remaining trials). All data were analysed in the same way as in Experiment 6.

6.3.2.1 Response Accuracy

Mean error rates across the conditions are illustrated in Figure 6.4 below. The response-accuracy model (marginal R^2 = .23, conditional R^2 = .53) revealed a significant main effect of Prime [$\chi^2(4) = 49.5$, $p < .001$]. Relative to unambiguous words (*M* = 9.6%, *SD* = 6.2), error rates were higher for homonyms (*M* = 28.7%, *SD* = 9.5; $p < .001$), irregular polysemes (*M* = 25.5%, *SD* = 6.7; $p < .001$), metaphors (*M* = 27.9%, *SD* = 7.4; $p < .001$), and metonyms (*M* = 18.1%, *SD* = 6.4; $p < .05$). Error rates for homonyms ($p < .01$), irregular polysemes ($p < .05$), and metaphors ($p < .05$) were also higher than those for metonyms. Further, there was a significant main effect of Target [$\chi^2(2) = 156.2$, $p < .001$]. Relatedness decisions were more erroneous to subordinate-sense (*M* = 38.5%, *SD* = 9.3) than dominant-sense (*M* = 21.4%, *SD* = 8.4; $p < .001$) and unrelated targets (*M* = 6.0%, *SD* = 5.8; $p < .001$), with a significant difference between the latter two ($p < .001$).

The response-accuracy model also revealed a significant Prime × Target interaction [$\chi^2(8) = 46.2$, $p < .001$]. For dominant-sense targets, planned contrasts indicated higher error rates for homonyms (*M* = 35.3%, *SD* = 13.6) than metaphors (*M* = 15.4%, *SD* = 9.0; $p < .01$), metonyms (*M* = 18.2%, *SD* = 9.6; $p < .05$), and unambiguous words (*M* = 10.5%, *SD* = 8.1; $p < .001$). Responses were also more erroneous to irregular polysemes (*M* = 27.4%, *SD* = 10.9) than unambiguous words ($p < .01$), while metaphors and metonyms did not differ from unambiguous control words. In contrast, for subordinate-sense targets, planned contrasts indicated higher error rates for metaphors (*M* = 63.6%, *SD* = 14.9) than homonyms (*M* = 43.0%, *SD* = 14.6; $p < .01$), irregular polysemes (*M* = 42.6%, *SD* = 10.5; $p < .01$), metonyms (*M* = 30.5%, *SD* = 9.9; $p < .001$), and unambiguous words (*M* = 13.1%, *SD* = 8.7; $p < .001$). Relatedness decisions were also more erroneous to homonyms ($p < .001$), irregular polysemes ($p < .001$), and metonyms ($p < .01$) than unambiguous words. For unrelated targets, there were no significant differences among the types of primes (all *ps* = 1).

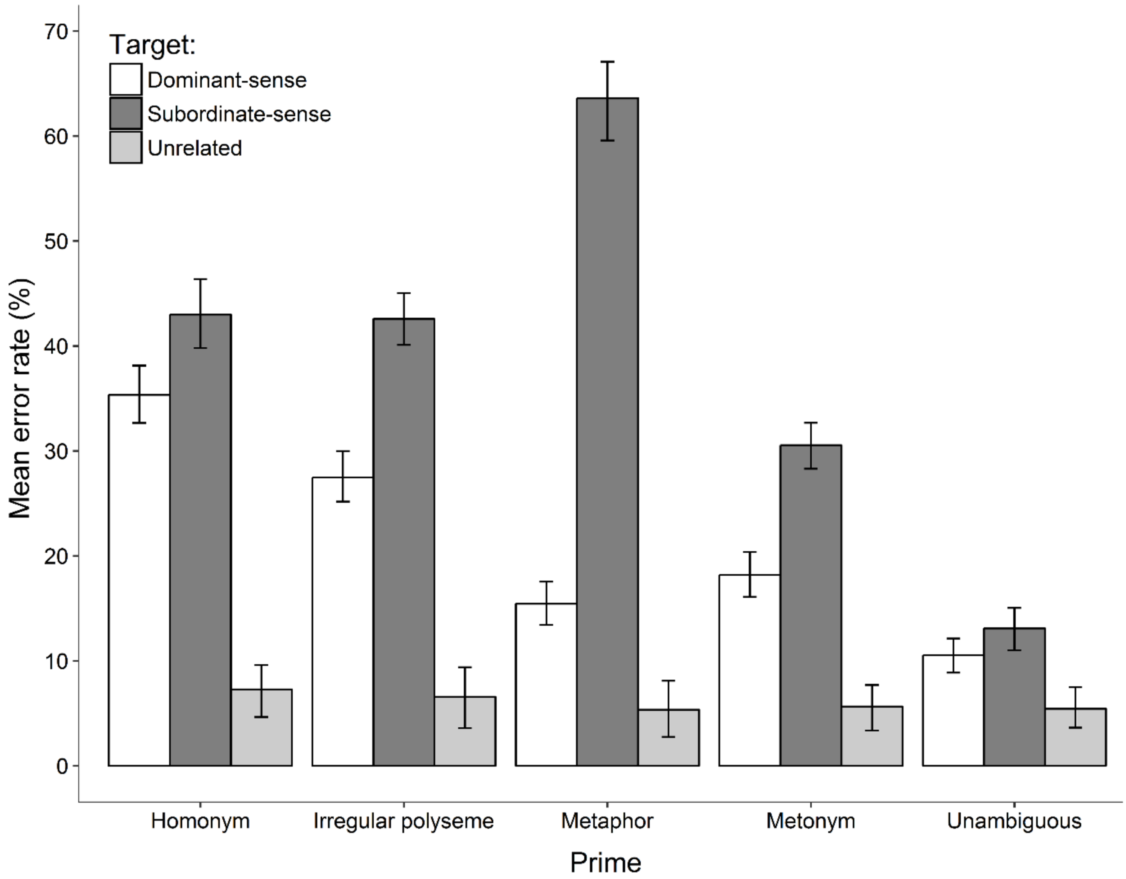

Figure 6.4: Experiment 7: Subject means of error rates across the conditions of Prime and Target. Error bars show 95% CIs adjusted to remove between-subjects variance.

6.3.2.2 Response Latency

Mean RTs across the conditions are illustrated in Figure 6.5 below. The response-latency model (marginal R^2 = 04, conditional R^2 = .47) revealed a significant main effect of Prime [$\chi^2(4) = 68.7$, $p < .001$]. Relative to unambiguous words (*M* = 624.0 ms, 95% CIs: 595.7, 653.9), RTs were higher for homonyms (*M* = 683.8 ms, 95% CIs: 652.4, 716.6; $p < .001$), irregular polysemes (*M* = 667.3 ms, 95% CIs: 636.6, 699.4; $p < .001$), metaphors (*M* = 668.2 ms, 95% CIs: 637.4, 700.3; $p < .001$), and metonyms (*M* = 641.7 ms, 95% CIs: 612.4, 672.5; $p < .05$). RTs for homonyms ($p < .001$), irregular polysemes ($p < .01$), as well as metaphors ($p < .01$) were also higher than those for metonyms. Further, there was a significant main effect of Target [$\chi^2(2) = 28.3$, $p < .001$]. Responses were slower to subordinate-sense (*M* = 669.9 ms, 95% CIs: 639.7, 701.5; $p < .001$) and unrelated (*M* = 662.8 ms, 95% CIs: 633.1, 693.7; $p < .001$) than dominant-sense targets (*M* = 637.7 ms, 95 %CIs: 609.1, 667.6), with no significant difference between the first two ($p = .23$).

The response-latency model also revealed a significant Prime × Target interaction [$\chi^2(8) = 46.3$, $p < .001$]. For dominant-sense targets, planned contrasts indicated higher RTs for homonyms (*M* = 666.8 ms, 95% CIs: 633.0, 702.4) than metaphors (*M* = 627.9 ms, 95% CIs: 596.5, 661.0; $p < .05$), metonyms (*M* = 632.4 ms, 95% CIs: 600.8, 665.7; $p < .05$), and unambiguous words (*M* = 602.3 ms, 95% CIs: 572.3, 633.9; $p < .001$). Responses were also slower to irregular polysemes (*M* = 661.2 ms, 95% CIs: 627.9, 696.1) than metaphors ($p < .05$) and unambiguous control words ($p < .001$). In contrast, for subordinate-sense targets, planned contrasts indicated higher RTs for metaphors (*M* = 731.1 ms, 95% CIs: 692.6, 771.8) than homonyms (*M* = 688.1 ms, 95% CIs: 653.0, 724.9; $p < .05$), irregular polysemes (*M* = 672.2 ms, 95% CIs: 637.5, 708.6; $p < .001$), metonyms (*M* = 643.1 ms, 95% CIs: 610.5, 677.3; $p < .001$), and unambiguous words (*M* = 620.6 ms, 95% CIs: 589.5, 653.1; $p < .001$). Responses were also slower to homonyms compared to metonyms ($p < .01$) and unambiguous words ($p < .001$), as well as irregular polysemes compared to unambiguous words ($p < .001$). For unrelated targets, contrasts indicated higher RTs for homonyms (*M* = 696.8 ms, 95% CIs: 662.1, 733.3) than metaphors (*M* = 649.8 ms, 95% CIs: 617.5, 683.9; $p < .001$), metonyms (*M* = 649.7 ms, 95% CIs: 617.3, 683.8; $p < .001$), and unambiguous words (*M* = 650.4 ms, 95% CIs: 617.0, 684.5; $p < .001$).

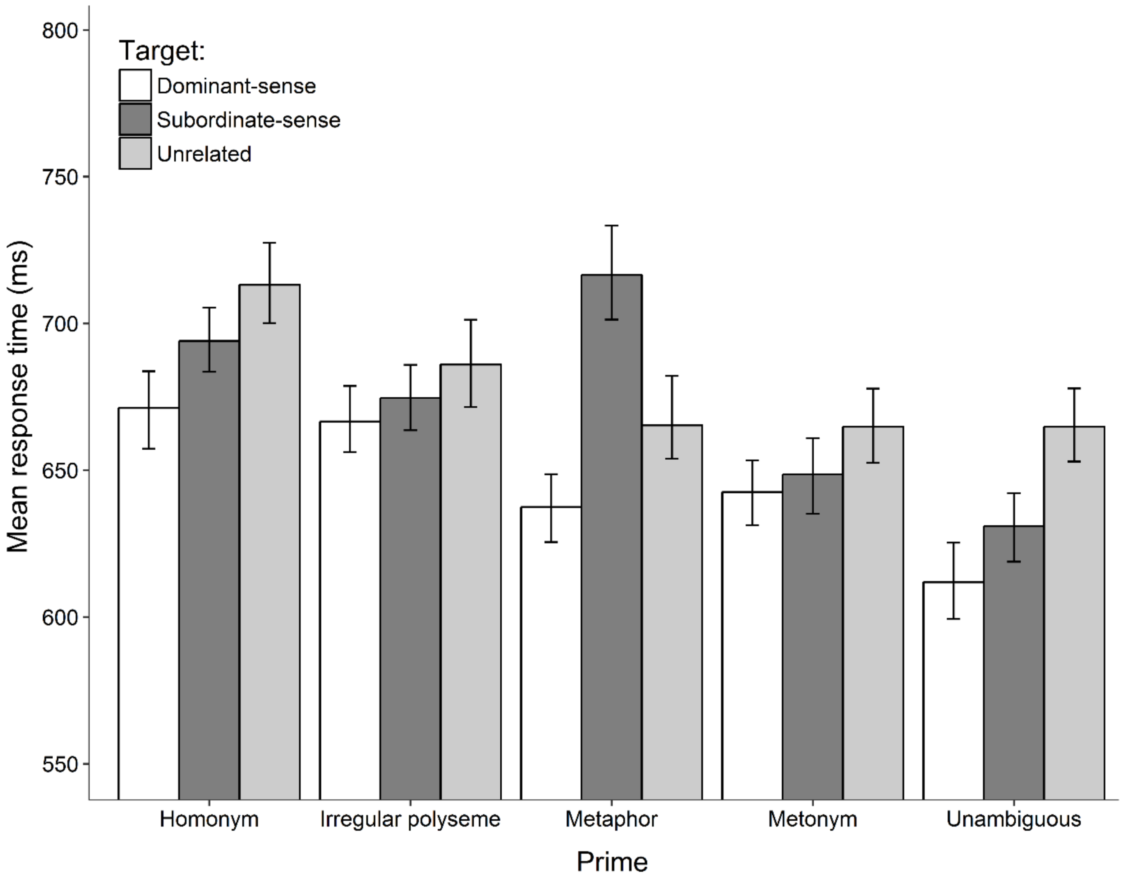

Figure 6.5: Experiment 7: Subject means of untransformed RTs across the conditions of Prime and Target. Error bars show 95% CIs adjusted to remove between-subjects variance.

As in Experiment 6, additional analysis explored whether the effects of Prime and Target on RTs were still significant when the lexical and semantic properties of the word stimuli were taken into account. Following the same procedures for covariate selection, the analysis included syntactic ambiguity and imageability at the prime level and imageability at the target level. All the effects remained significant, except for slower responses to homonyms and irregular polysemes than metonyms across the three types of trials, slower responses to homonyms than metonyms on dominant-sense trials, slower responses to irregular polysemes than metaphors on dominant-sense trials, and slower responses to metaphors than homonyms on subordinate-sense trials (see Figure 6.5 above) that were no longer significant.

6.3.2.3 Experiments 6 & 7 Combined: Response Accuracy

The following analyses contrasted task performance in Experiments 6 and 7 to examine the influence of prime-word duration (200 ms in Experiment 6, 700 ms in Experiment 7) on the processing of ambiguous words. In particular, the key question raised in the introduction to Experiment 7 was whether the longer prime duration would provide participants with more time to retrieve and select the subordinate sense of metaphorical polysemes. The analyses were the same as those conducted for each experiment separately, except that they included the additional factor of Experiment (6 vs. 7). All models included random intercepts for subjects and items and a random by-subjects slope for Experiment. Planned contrasts (with the Holm-Bonferroni correction) explored between-experiments differences in the previously reported effects of Prime and Target.

The response-accuracy model (marginal R^2 = .21, conditional R^2 = .51) showed a significant Experiment × Target interaction [$\chi^2(2) = 20.8$, $p < .001$]. Although there were far more errors to related than unrelated targets in both experiments, this effect was smaller in Experiment 6 than Experiment 7, both for dominant-sense (Experiment 6: 5.4% difference, relative to unrelated targets; Experiment 7: 13.4% difference; $p < .01$) and subordinate-sense targets (Experiment 6: 28.2% difference; Experiment 7: 32.5% difference; $p < .001$). Further, there was a significant Experiment × Target × Prime interaction [$\chi^2(8) =$ 16.7, $p < .05$]. Pairwise comparison analysis revealed that the aforementioned effect was restricted to balanced homonyms, both for dominant-sense (Experiment 6: 22.7% difference, relative to unrelated targets; Experiment 7: 28.1% difference; $p < .05$) and subordinate-sense targets (Experiment 6: 31.1% difference; Experiment 7: 35.8% difference; $p < .05$). This shows that participants were better (i.e., made fewer errors) at comprehending the two meanings of balanced homonyms when the words were presented for a shorter period of time.

6.3.2.4 Experiments 6 & 7 Combined: Response Latency

The response-latency model (marginal $R^2 = .06$, conditional $R^2 = .45$) revealed a significant Experiment × Prime interaction [$\chi^2(4) = 11.0$, $p < .05$] that was due to greater disadvantage effects for homonyms ($p < .05$) and irregular polysemes ($p < .01$) in Experiment 6 (homonyms: 71.4 ms difference, relative to unambiguous control words; irregular polysemes: 60.2 ms difference) than Experiment 7 (homonyms: 59.8 ms difference; irregular polysemes: 43.3 ms difference). There was also a significant Experiment × Target interaction [$\chi^2(2) = 58.3$, $p < .001$]. Although responses were slower to unrelated than related targets in both experiments, the effect was greater in Experiment 6 than Experiment 7, both relative to dominant- (Experiment 6: 50.2 ms difference; Experiment 7: 25.1 ms difference; $p < .001$) and subordinate-sense targets (Experiment 6: 15.4 ms difference; Experiment 7: -7.1 ms difference; $p < .001$).These results remained significant in the analysis with relevant covariates (i.e., syntactic ambiguity and imageability at the prime level, imageability at the target level).

6.3.3 Discussion

As in Experiment 6, an overall disadvantage effect appeared only for balanced homonyms and irregular polysemes. Between-experiments analyses revealed that the manipulation of prime duration had an impact on responses to metonyms but not metaphors. The cost associated with metonyms in the subordinate sense did not arise in Experiment 7, suggesting that the additional process of retrieving that sense, which is assumed to underlie the cost in Experiment 6, was completed within the additional 500 ms. This indicates that even though it may take more time to activate the subordinate sense, metonymy does not produce competition even when the dominant and the subordinate sense have been activated to the same extent. In contrast, the cost associated with metaphors in the subordinate sense arose again and was comparable in both experiments. It appears, then, that metaphor does not produce competition because language users do not retrieve the subordinate sense out of context, even when there is enough time do so (see also Klepousniotou & Baum, 2005a).

6.4 General Discussion

The work reported in this chapter is the first to examine how polysemes are comprehended in the absence of contextual bias, making a significant contribution to the current literature that has to date only explored how these words are recognised (e.g., Klepousniotou & Baum, 2007; Rodd et al., 2002) or processed within sentences (e.g., Frazier & Rayner, 1990; Frisson & Pickering, 1999) and phrases (e.g., Klein & Murphy, 2001; Klepousniotou et al., 2008).

Indeed, using the disadvantage effect as an index of semantic competition in the present experiments revealed important differences in how irregular, metonymic, and metaphorical polysemes are processed and, by inference, represented in the mental lexicon. These findings and their implications for the ambiguity literature are discussed in turn.

6.4.1 Irregular Polysemy

Experiments 6 and 7 found an equivalent processing disadvantage for irregular polysemes and balanced homonyms, in particular on trials instantiating the dominant or the alternative sense of these words. This novel finding suggests that the two forms of ambiguity may be processed in a very similar manner after all. Even though the senses of irregular polysemes are related to some extent with respect to function (e.g., “restaurant/website menu”), physical properties (e.g., “plastic/wheat straw”), or other conceptual features, it appears that these words still have multiple semantic representations that compete for activation and slow word comprehension, just like the unrelated meanings of homonyms.

This interpretation receives further support from Experiment 5, in which learning new senses of irregular polysemes was demonstrated to give rise to competition with existing senses of these words. It is important to mention, however, that although the evidence from processing existing ambiguous words (Experiments 6 & 7) and learning new meanings for familiar words (Experiment 5) indicates competition between multiple semantic representations, there seems to be slight disagreement on the precise nature of the representations. The finding of comparable disadvantage effects for irregular polysemes and homonyms in Experiments 6 and 7 suggests that irregular polysemes have separate representations. The finding of a significantly smaller effect for irregular polysemes than homonyms in Experiment 5, on the contrary, suggests that irregular polysemes have separate but overlapping representations that do not compete as much as the separate representations of homonyms (for a detailed discussion, see 5.4.1 Meaning Relatedness in Chapter 5). There are several significant differences between Experiment 5 and Experiments 6 and 7 that make it difficult to unravel this inconsistency. One possibility, for example, is that the findings of the experiments are not comparable, in that the influence of newly-formed representations for fictitious meanings, especially in the early stages of meaning consolidation, differs from that of robust representations for well-known meanings. Whether irregular polysemes have completely separate representations that compete for activation of all their features or overlapping representations that compete only for activation of their unique features remains to be established. What is clear, nevertheless, from all three experiments, is that irregular polysemes have more than one representation at the semantic level.

This separate-representations account for irregular polysemes is in line with the findings of Rabagliati and Snedeker (2013). In their study, participants named aloud pictures illustrating one of the word referents of homonyms (e.g., “scary/baseball bat”), irregular polysemes (e.g., “shirt/emergency button”), and metonyms (“stalks/kernels of corn”) that were accompanied by same-name foils, or pictures illustrating the other word referent. The study found that participants produced significantly more ambiguous bare labels (e.g., “button” instead of “emergency button”) for irregular polysemes than metonyms, at similar levels as for homonyms. Rabagliati and Snedeker (2013) argued that speakers fail to avoid ambiguity when naming irregular polysemes, and detect it only after the fact, because the different senses of the words are accessed from separate semantic representations, just like the meanings of homonyms. Thus, there is now evidence from both comprehension and production paradigms to show that understanding irregular polysemes and homonyms involves very similar representations and processing mechanisms.

The present findings present a major challenge for the shared features account developed by Brocher et al. (2016). According to this account, irregular polysemes have a single representation comprising the shared and unshared features of the different word referents. Readers first access an overall meaning of an irregular polyseme based on activation of the shared features, and later disambiguate the word toward a specific sense, usually the dominant one, based on activation of the unshared features. The only condition in which the account predicts any processing cost is the one in which context preceding an irregular polyseme supports its subordinate sense, leading to suppression of the unshared features corresponding to the dominant sense that are activated whenever the word is encountered. The present experiments did not create such a condition, and yet a significant processing cost was observed. Irregular polysemes appeared in isolation and slowed responses regardless of whether they were later followed by dominant-sense, subordinate-sense, or unrelated targets. Overall, then, the evidence presented in this chapter is incompatible with Brocher et al.'s (2016, 2018) account, or any accounts that posit shared representations for the senses of irregular polysemes. Such accounts are unable to explain why words that are supposed to have only one representation would produce competition and behave like words with multiple representations (i.e., homonyms).

6.4.2 Metonymic Polysemy

Turning to metonyms, Experiments 6 and 7 found that these words do not show a typical disadvantage effect. Metonyms incurred a small processing cost only when disambiguated toward the subordinate sense (in particular, in Experiment 6 with the short prime duration). It seems that metonyms escape competition because they have only one representation, and therefore no competitors in the race for semantic activation.

The finding of the slowing for metonyms in the alternative sense has particularly important implications for the existing work on the representation and processing of these words, especially the core representation (e.g., Klepousniotou et al., 2008; Pustejovsky, 1995) and the underspecification account (e.g., Frisson, 2009; Nunberg, 1979). While both accounts posit a single representation, they disagree on what such a representation comprises, and how readers settle on a specific interpretation of a metonym. Klepousniotou et al. (2008) proposed that metonyms have a core representation that generally corresponds to the features of the dominant sense. Readers access that sense from the representation whenever a metonym is encountered, but then derive the alternative sense via a productive rule. Thus, under the core representation account, there should be a slight delay in access to the alternative sense. In contrast, Frisson and Pickering (1999) argued that readers first access a general meaning of a metonym from an abstract, underspecified representation, and then home in on a specific sense of the word using surrounding context. Under their underspecification account, context should help to flesh out both the dominant and the alternative sense without noticeable difficulty.

The present findings are evidently consistent with the predictions of the core representation account (e.g., Copestake & Briscoe, 1995; Klepousniotou et al., 2008; Pustejovsky, 1995). Participants did find it more difficult to process the subordinate sense of metonyms even when presented with supporting context, or in this case the disambiguating target word (see also Foraker & Murphy, 2012; Lowder & Gordon, 2013). Critically, this effect was observed when the ambiguous words were presented for 200 ms (Experiment 6) but not for 700 ms (Experiment 7), which suggests that the difficulty with the subordinate sense may arise only in the early stages of processing. While evidence from semantic priming found the different senses of metonyms to be activated to the same extent at short and long SOAs (Klepousniotou & Baum, 2005; Klepousniotou et al., 2012; MacGregor et al., 2015), the current evidence from relatedness decisions is the first to show the seemingly brief moment when this is not the case. Although all senses ultimately reach equivalent levels of activation, there is a slight delay for subordinate senses at first, most likely because they are not accessed from the lexicon but must be derived from the dominant sense via a productive rule.

Taken together, the present investigation lends support to the prevalent view that the highly related senses of metonyms are stored within a single representation. Although understanding metonyms in minimal context does not involve competition, there seems to be a small and transient cost associated with processing the alternative sense of these words. This novel finding helps to tease apart conflicting accounts of metonym representation. It corroborates the core representation variant, in particular that exemplified by Klepousniotou et al.

(2008), according to which the dominant sense is fully specified within the representation, whereas the subordinate sense must undergo the additional process of on-line derivation.

6.4.3 Metaphorical Polysemy

Experiments 6 and 7 found no evidence of competition between the dominant/literal and the subordinate/figurative sense of metaphors, most likely due to very weak activation of the latter in minimal context. The results showed that participants rarely retrieved the subordinate sense, regardless of how long the ambiguous words were presented for, and that a substantial processing cost arose only when they did. In other words, metaphor slowed comprehension only when the reading of a target word instantiating the unexpected figurative sense made participants engage in evidently effortful and time-consuming retrieval of that sense.

The view that readers do not fully retrieve the subordinate sense of metaphors out of context is consistent with the results of the stimulus rating studies conducted prior to Experiment 6 (see 6.2.1.2 Stimuli in this chapter). These studies revealed that raters failed to detect the semantic relatedness between most of the metaphors and their subordinate-sense targets, unless their interpretation of the words was primed by strong sentential context supporting this particular sense. The view is also consistent with semantic priming studies that showed evidence of subordinate-sense activation for metaphors in biasing (Klepousniotou, 2002; Klepousniotou & Baum, 2005b) but not in neutral/minimal context (Klepousniotou & Baum, 2005a; Klepousniotou, Gracco, & Pike, 2007;

but cf. MacGregor et al., 2015). Overall, then, it appears that naturalistic and elaborate context is needed to arrive at the figurative interpretation of metaphors in both offline and online tasks.

With regards to representational issues, the present findings can be interpreted in favour of either the separate- or the single-representation account. One possibility is that metaphors have separate representations but escape semantic competition simply because the subordinate sense is not sufficiently activated to compete with the dominant one. In other words, the loosely related senses of metaphors may be represented and processed like the unrelated meanings of unbalanced homonyms (see 3.4.1 Unbalanced Homonymy in Chapter 3). Another possibility is that metaphors do not show any effects of competition because they have only one representation for the dominant sense. In this respect, metaphors are represented and processed like metonyms, with the dominant sense accessed from the core representation and the subordinate one generated on-line. Note, however, that although the separate-representations account holds true for the current findings, this is not the case for other studies on the processing of metaphors (Klepousniotou, 2002; Klepousniotou & Baum, 2005a; 2007; Klepousniotou et al., 2007, 2012; MacGregor et al., 2015). As discussed earlier, the studies have consistently shown that having literal and figurative senses facilitates rather than inhibits word processing, rendering the separate-representations account unable to explain metaphor effects in lexical decision and semantic priming paradigms.

To summarise, the findings tend to support the idea that the figurative sense of metaphors is neither stored in the mental lexicon nor retrieved without contextual support. Language users may derive the figurative sense from the literal one by applying sense extension rules, or construct it based on surrounding context. The former is particularly likely for “regular” metaphors, (such as those used in the present experiments) that are characterised by conventional and relatively predictable patterns of sense extension (e.g., animal for human trait and body part for object alternations). Regardless of how the subordinate sense of metaphors is generated, both Experiments 6 and 7 demonstrate this process to be more effortful and slower than that for metonyms. This could be because the representations of metaphors, highly specialised for the dominant sense, contain very few, if any, features that are shared by the subordinate sense.

6.4.4 Further Implications

The findings of Experiments 6 and 7 place serious constraints on PDP models that distinguished between homonymy and polysemy (Armstrong & Plaut, 2008; Rodd et al., 2004). In the current form, the models assume that polysemes have separate but overlapping representations, which results in either minimal (Armstrong & Plaut, 2008) or moderate semantic competition (Rodd et al., 2004). Experiment 6 and 7 demonstrate that this assumption is incorrect, or at best, that it does not apply to all types of polysemes. Although the models can readily explain the disadvantage effect for irregular polysemes, the finding of no such effect for both metaphors and metonyms is another matter. These words do not show any signs of competition; they appear to have a single representation comprising the features of the dominant sense, some of which happen to be also true of the subordinate sense. The implication is not that we should abandon PDP models altogether, but that they need to adapt their assumptions in line with the present as well as earlier findings on the processing of metaphor and metonymy. This is discussed in more detail in Chapter 7.

The work discussed in this chapter is also of great value to the broader literature on ambiguity that has, thus far, largely overlooked the distinctions within polysemy. The findings demonstrate that the ambiguity between multiple related senses is not a uniform phenomenon. Irregular polysemes behave like homonyms, in that they also show effects of competition between separate semantic representations. Metaphors and metonyms, on the other hand, may both have only one representation, but there is a striking difference in how these words are processed when the subordinate sense is required. This indicates that the type of sense extension is perhaps the most important property to take into consideration when exploring the processing of polysemy. If the impact of polysemy were primarily driven by the number of related senses, as suggested by studies of lexicographic polysemy (e.g., Armstrong & Plaut, 2011; Jager & Cleland, 2016; Rodd et al., 2002), the present experiments would have found similar patterns of responses for irregular polysemes, metaphors, and metonyms that were matched for this variable. Likewise, if the impact were driven by sense relatedness, as assumed by studies of subjective polysemy (e.g., Haro et al., 2007; Hino et al., 2006; Klein & Murphy, 2001), there would not have been any differences in the processing of irregular polysemes and metaphors that were given similar ratings of relatedness. The implication is not that sense count and sense relatedness are irrelevant, but that their effects may be observable within each form of polysemy, such that, for example, it is more difficult to comprehend many-sense than few-sense irregular polysemes. This idea, however, remains a speculation until future studies manipulate or vary sense count and sense relatedness over and above the linguistic distinction into irregular, metaphorical, and metonymic polysemy.

6.4.5 Conclusions

To summarise, this work provides a novel type of evidence, from relatedness decisions, of representational and processing differences within polysemy. The results corroborate the separate-representations account for irregular polysemes (Experiment 5 in this thesis; Rabagliati & Snedeker, 2013), similar to homonyms, and the single-representation account for metaphors and metonyms (e.g., Frazier & Rayner, 1990; Klepousniotou & Baum, 2007; MacGregor et al., 2015). The results also help to differentiate between existing variants of the account for metonyms, demonstrating, for the first time, that the subordinate sense of these words is derived from the dominant sense (Klepousniotou et al., 2008), rather than surrounding context (Frisson, 2009). Most importantly, this work reveals that not all forms of ambiguity produce semantic competition in word comprehension. PDP models, such as those proposed by Armstrong and Plaut (2008) and Rodd et al. (2004), are currently unable to explain why this is the case, and need to revisit their assumptions on the nature of polysemy, metaphor and metonymy in particular, if they are to remain relevant in the field.

Chapter 7: General Discussion

One way we can use language to communicate an infinite number of ideas through a finite number of words is by compressing emerging meanings into existing words, such that, over time, most of the words become associated with multiple different meanings in different contexts. The ability to select a single, contextually appropriate meaning without being overtly distracted by the myriad of other possible meanings is, therefore, a crucial component of any theory of language comprehension. Indeed, the importance of understanding how word meanings are stored in the mental lexicon and selectively accessed is highlighted by the extensive psycholinguistic literature dedicated to these issues over the last two decades (e.g., Armstrong & Plaut, 2008; Brown, 2008; Cai et al., 2017; Hino et al., 2010; Kandhadai & Federmeirer, 2007; Klepousniotou, & Baum, 2007; MacGregor et al., 2015; Pickering & Frisson, 2001; Rabagliati & Snedeker, 2013; Rayner et al., 2006; Rodd et al., 2002; Sereno et al., 2006).

One under-researched finding in past literature is the slower response and reading times for ambiguous versus unambiguous words in tasks that require meaning selection in neutral/minimal context (Duffy et al., 1988; Frazier & Rayner, 1990; Gottlob et al., 1990; Hino et al., 2006; Hoffman & Woollams, 2015; Jager & Cleland, 2015; Pexman et al., 2004; Piercey & Joordens, 2000; Rayner & Duffy, 1986). Although this ambiguity disadvantage appears to be established, remarkably little is known about its locus. Some researchers have taken the effect as evidence of semantic competition in ambiguity processing, in line with simulations of PDP models (e.g., Armstrong & Plaut, 2008; Joordens & Besner, 1993; Kawamoto, 1993; Rodd et al., 2004). These models assume that the distinct meanings or senses of ambiguous words have separate semantic representations that compete for activation of their respective features, and that it takes more time to comprehend ambiguous words because this competition delays reaching a stable, consistent pattern of semantic activation. In contrast, others have explained the effect in terms of ambiguity-specific difficulties in response selection (Hargreaves et al., 2011; Hino et al., 2006; Pexman et al., 2004; Siakaluk et al., 2007). In particular, Pexman et al. (2004) suggested that ambiguity does not slow semantic activation, but places additional demands on the decision-making system when, for example, the different meanings of words are inconsistent with a single yes-no response.

The work reported in this thesis focussed on the ambiguity disadvantage in single-word processing. In all the experiments, participants made relatedness decisions to ambiguous and unambiguous primes followed by targets that supported one of their meanings or were unrelated. The semantic competition and decision-making accounts were explicitly tested by keeping constant the decision-making demands of the tasks while manipulating factors that impact on the semantic activation process – namely, meaning frequency (Experiments 2-4) and meaning relatedness (Experiments 5-7). These manipulations were successful in uncovering which types of ambiguity produce the disadvantage effect and why.

7.1 Summary of the Findings

Chapters 2-4 focussed on homonymy, under the expectation that if any form of ambiguity entailed semantic competition, it would be primarily observed for words that have multiple unrelated meanings. Experiments 2-4 contrasted homonyms with balanced and unbalanced meaning frequencies which were selected through extensive norming conducted as part of Experiment 1. The results showed that the ambiguity disadvantage is of a semantic nature as it is sensitive to meaning frequency and its well-documented impact on semantic activation (e.g., Binder & Rayner, 1998; Chen & Boland, 2008; Klepousniotou et al., 2012; Rayner & Frazier, 1989; Simpson & Burgess, 1985; Swaab et al., 2003). The disadvantage effect was observed for balanced but not unbalanced homonyms, extending earlier findings from sentence reading (Duffy et al., 1988; Rayner & Duffy, 1986) to single-word processing.

The findings are consistent with the predictions of PDP models (e.g., Joordens & Besner, 1994; Rodd et al., 2004), especially those that account for the impact of meaning frequency (Kawamoto, 1993). Homonymy slows word comprehension due to competition at the semantic level, albeit the degree of this competition depends on meaning frequency. For unbalanced homonyms, meaning frequency biases the activation process in favour of the HF meaning, such that the representation of the LF counterpart does not reach sufficient activation to engage in competition. For balanced homonyms, on the other hand, the representations are activated in parallel and to a similar extent, and must therefore compete for further activation.

The findings of the current experiments challenge the alternative account which suggests that the ambiguity disadvantage is due to decision-making difficulties (Hargreaves et al., 2011; Hino et al., 20006; Pexman et al., 2004; Siakaluk et al., 2007). To begin with, Experiments 2 and 3 (Chapter 3) found a significant slowing for balanced homonyms on related trials where the different meanings triggered conflicting responses to the target as well as on unrelated trials where they did not. This is incompatible with Pexman et al.'s (2004) proposal that ambiguity slows relatedness decisions exclusively on related trials, when it creates response conflict. Furthermore, Experiment 4 (Chapter 4) revealed that the disadvantage effect for balanced homonyms arises during semantic activation when processing the ambiguous word, and that it likely reflects increased activity in the LIFG which is assumed to resolve competition between multiple activated representations (e.g., Bedny et al., 2008; Bilenko et al., 2009; Novick et al., 2009). This finding, in particular, temporally places the disadvantage effect during the semantic activation process, and thus refutes any proposal that the ambiguity disadvantage arises later and is due to decision making during response selection.

Chapters 5 and 6 focussed on polysemy. The main aim was to develop a comprehensive account of the ambiguity disadvantage and determine whether PDP models (e.g., Borowsky & Masson, 1996; Kawamoto et al., 1994; Rodd et al., 2004) are correct to predict semantic competition for all forms of ambiguity. Experiment 5 (Chapter 5) probed the impact of irregular polysemy by examining how learning new loosely related meanings for previously unambiguous words affects readers' ability to comprehend their existing meanings. Experiments 6 and 7 (Chapter 6) then explored the processing of irregular polysemy as well as metaphor and metonymy in natural language. Collectively, the results revealed that semantic competition arises only for irregular polysemes, which are seemingly represented and processed like homonyms (see also Rabagliati & Snedeker, 2013). Metaphorical and metonymic polysemes do not show any signs of competition, most likely because they have only one representation in semantic space, as also suggested by evidence from other paradigms (e.g., Frazier & Rayner, 1990; Frisson & Frazier, 2005; Klepousniotou & Baum, 2007; Klepousniotou et al., 2008; Lopukhina et al., 2018; MacGregor et al., 2015; McElree et al., 2006). Overall, then, the findings in this thesis indicate that the prediction of competition between multiple semantic representations in current PDP models does not hold true for all forms of ambiguity. Rather, competition seems to arise primarily for words with multiple meanings that have relatively equal frequency and loose, if any, semantic connections between them.

7.2 Implications and Future Directions

The present work shows that although PDP models correctly explain the ambiguity disadvantage in terms of semantic competition, they need to revisit their assumptions about the representation and processing of polysemy. One way to account for the findings for metaphors and metonyms (Experiments 6 & 7) is to assume that their orthographic pattern of activation is associated with a single semantic pattern that largely maps onto the dominant sense. Subordinate senses are not stored directly in the lexicon; instead, rules of sense extension are stored that readers can apply to derive those senses on-line (see also Copestake & Briscoe, 1995; Klepousniotou et al., 2008). If this is true, PDP models that implement single form-to-meaning mappings for metaphors and metonyms and provide ways/rules to construct subordinate senses on-line in supporting context should have no difficulty simulating behavioural data and explaining why these words avoid semantic competition.

Developing such models is certainly not an easy task, but it has several benefits to the field. In particular, although there is overwhelming agreement in the linguistic and psycholinguistic literature that metonyms have only one representation that acts as a gateway for their multiple senses, remarkably little attention has been given to specifying what information actually comprises such a representation. A number of proposals have been put forward, though they have been vague and strictly theoretical in nature. The proposals have ranged from the idea of an impoverished/underspecified semantic representation that corresponds to an overall, abstract meaning (e.g., Frisson & Pickering, 1999) to the idea of a rich semantic representation that corresponds to a common core, encompassing features that are true of all the senses (e.g., Klepousniotou et al., 2008). Working towards computational instantiations of these ideas will require spelling out which information is stored in the lexicon and which information is not, and will consequently provide a way to distinguish between these proposals and test them.

With regards to the issue of lexical-semantic representation in a wider context, the evidence presented in this thesis is most compatible with the view that our knowledge of words and their meanings is distributed over a number of neuron-like units corresponding to orthographic and semantic features (e.g., Gaskell & Marslen-Williams, 1997; Harm & Seidenberg, 1999, 2004; Rumelhart, Hinton, & McClelland, 1986; Seidenberg & McClelland, 1989). Simply put, PDP models, which take this "distributed" view, assume that ambiguous words have a single representation at the orthographic level but multiple representations at the semantic level, which naturally gives rise to competition during the semantic activation process (e.g., Joordens & Besner, 1994; Kawamoto, 1993; Rodd et al., 2004). PDP models can, therefore, easily explain why it takes more time to comprehend (some) ambiguous versus unambiguous words. In fact, they seem to be well-placed to accommodate a number of findings, including the effects of context (Twilley & Dixon, 2000) and meaning frequency in ambiguity processing (Experiments 2-4; Kawamoto, 1993; Rodd et al., 2016) as well as the effects of ambiguity in word recognition (Armstrong & Plaut, 2008; Rodd et al., 2004), naming (Kawamoto, 1993), word learning (Experiment 5; Rodd et al., 2012), reading aloud (Rodd, 2004), and word comprehension (Experiments 2-7).

In contrast, the finding of the ambiguity disadvantage is incompatible with the “localist” view, according to which the mental lexicon contains higher-order units corresponding to words, rather than lower-order units corresponding to their orthographic or semantic features (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1994; Levelt, Roelofs, & Meyer, 1999; Morton, 1969; Rumelhart & McClelland, 1982). Localist models (Jastrzembski, 1981; Rubenstein et al., 1970) propose that each meaning of an ambiguous word has its own orthographic and semantic representation - a dedicated entry in the lexicon. Activation of the entries is assumed to arise independently, such that meaning selection terminates whenever one of them reaches a sufficient level of activation. Thus, localist models do not predict ambiguity to inhibit word processing at all; difficulties may appear only when readers reinterpret a word after an incorrect meaning has been selected (but see Gottlob et al., 1999).

The proposal that the ambiguity disadvantage arises during contextual integration is, however, problematic. For example, it stands in stark contrast to the finding of the disadvantage during the processing of the ambiguous word itself, rather than during the processing of disambiguating context (Experiment 4 in this thesis; Duffy et al., 1988; Frazier & Rayner, 1990; Rayner & Duffy, 1986). It seems that while the localist framework remains influential in the field of word recognition (for a review, see Coltheart, 2006), it has reached an impasse in the field of ambiguity resolution as it cannot accommodate ambiguity effects in word comprehension. In particular, Experiments 2-7 showed that these effects are underlain by semantic activation, not integration, processes.

It is important to note that the present work has implications that go beyond the ambiguity literature. Ambiguity provides us with a ubiquitous and natural manipulation of semantics that can be used to uncover the basic principles underlying language and cognition in general. There is already an impressive number of studies that have used ambiguous words as the means for examining hemispheric asymmetries in language processing (e.g., Faust & Lavidor, 2003; Klepousniotou et al., 2014); contextual constraints on lexical access (e.g., Onifer & Swinney, 1981; Vu, Kellas, Metcalf, & Herman, 2000); inhibition mechanisms in individuals with schizophrenia (e.g., Salisbury, 2010; Titone, Levy, & Holzman, 2000), autism (e.g., Hala, Pexman, & Glenwright, 2007; Norbury, 2005), and brain damage (e.g., Hagoort, 1993; Klepousniotou & Baum, 2005b); as well as acquisition of conceptual knowledge (e.g., Srinivasan & Snedeker, 2011, 2013). However, in order to use ambiguity as a window into such processes, we must fully understand how ambiguity operates in the first place. We must recognise, for instance, that homonymy can hinder word comprehension due to either semantic competition or effortful retrieval and integration of the highly uncommon meaning (Experiments 2 & 3; Rayner & Duffy, 1986); that it is easier to learn new related than unrelated meanings for words (Experiment 5; Rodd et al., 2012); and that context plays a particularly large role in disambiguating unbalanced homonyms (Experiments 2-4) and metaphors (Experiments 6 & 7; Davies et al., 2017; Klepousniotou & Baum, 2005a). Only then can we make the most of ambiguity as a tool to examine, for example, how the right and the left hemisphere contribute to the processing of meaning.

Understanding how the language system deals with ambiguous words is also important if we are to make further progress in the field of reading. Existing models of skilled, adult reading focus exclusively on orthography and phonology and often propose an unclear or reductionist view for the role of semantics. For instance, the dual route cascaded model (Coltheart et al., 2001) is, to date, one of the most influential models of visual word recognition and reading aloud. However, while the model has been studied quite extensively, its semantic component has yet to be conceptualised and implemented. Our knowledge of word meanings plays a large role in learning and reading words (for reviews, see Balota, Yap, & Cortese, 2006; Pexman, 2012; Taylor, Duff, Woollams, Monaghan, & Ricketts, 2015), which comes as no surprise when it is recognised that the purpose of words is not to be simply recognised but to convey meaning. Although there are many semantic influences on word processing that models of reading need to account for, the greatest challenge may lie in recognising that over 80% of the words we use are in some way ambiguous, and that, broadly speaking, these words facilitate early recognition processes (e.g., Hsiao & Nation, 2018; Klepousniotou & Baum, 2007; Rodd et al., 2002; Taler, Kousaie, & López Zunini, 2013) but complicate later comprehension processes (e.g., Experiments 2-7; Hoffman & Woollams, 2015; Kambe, Rayner, & Duffy, 2001; Klepousniotou et al., 2008; Rodd, Johnsrude, & Davis, 2010). Thus, there is a clear need for models of reading to incorporate a semantic component, one that would resolve ambiguity in the way that language users do.

This issue naturally extends to the field of reading development, where much attention has focussed on examining difficulties in word recognition but not in word comprehension (see Castles, Rastle, & Nation, 2018). Although cracking the alphabetic code is the very first step towards learning to read, we must not forget that the ultimate goal for children is to develop higher-level skills that will allow them to construct meaning. This is illustrated by the first sentence in a recent reading comprehension test for 10-/11-year-old children in the final year of primary school in the UK (Standards & Testing Agency, 2016) – "Dawn was casting spun-gold threads across a rosy sky over Sawubona game reserve". To correctly understand the sentence, children must not only recognise each of the words, but also select their meanings, appropriate for the context. They need to work out, for example, that "Dawn" does not refer to a girl's name, and that "game" does not refer to a type of sport. Therefore, there seems to be an expectation that children will have learnt to infer meaning by the end of primary school.

Evidence suggests that this is not the case, at least for some typically developing children who may read fluently but have difficulty in understanding words with multiple meanings (Booth, Harasaki, & Burman, 2006; Hala et al., 2007; Henderson, Snowling, & Clarke, 2013; Khanna & Boland, 2010; Simpson & Foster, 1986; van der Schoot, Vasbinder, Horsley, Reijntjes, van Lieshout, 2009). The work presented in this thesis does not explain why these children struggle, or how to improve their comprehension, but it does provide important information for future research attempting to do so. In particular, the present studies show that even skilled adult readers often misinterpret words in minimal context due to strong bias toward the dominant meaning (Experiments 2-4) or sense of the words (Experiment 6 & 7). This suggests that using context, and perhaps other cues, to override this bias may be one of the key higher-level language skills that children need to master before entering secondary school (see also Nievas & Justicia, 2004; Rabagliati, Pylkkänen, & Marcus, 2013).

With regards to methodological implications, two contributions deserve special attention. First, Experiment 1 provides the first UK-based norms of meaning frequency for a large set of homonyms that were carefully selected and validated with future studies in mind. The norms are of particular value to UK-based researchers, and it is recommended that they are used for two main reasons. To begin with, Experiments 2-4 demonstrated, for the first time, that balanced and unbalanced homonyms differ in how their meanings are activated out of context and how they affect word comprehension, which highlights the need to either manipulate or control for meaning frequency when investigating homonymy processing. Furthermore, Experiment 1 revealed that estimates of meaning frequency vary between speakers of the English language depending on their dialect. This finding, which has already been cited and taken into consideration by recent studies (Gilbert, Betts, Jose, & Rodd, 2017; Rice, Beekhuizen, Dubrovsky, Stevenson, & Armstrong, 2018), highlights the need for UK-based norms when testing UK-based participants.

The second methodological contribution, demonstrated in Experiments 6 and 7, is the evidence that the ambiguity between multiple related senses is a complex phenomenon. Irregular polysemes behave like homonyms, in that they also show effects of competition between multiple representations in semantic space. Metaphors and metonyms, on the other hand, may both have only one representation, but there is a striking difference in how the words are processed when the subordinate sense is required. This raises the possibility that the handful of studies which failed to find a relatedness effect (Haro et al., 2017; Hino et al., 2006; Klein & Murphy, 2001; Lin & Ahrens, 2010) may have done so because they used irregular polysemes that are evidently processed like homonyms. The implication is that in order to explore polysemy, one needs to clearly specify what is meant by this term. Defining polysemy solely based on the number of senses in a dictionary or subjective ratings of sense relatedness will not suffice. Rather, researchers should focus on the nature of the sense extension, as evidence shows that having irregular, figurative, or regular sense extensions clearly makes a difference, both in terms of representation and processing.

7.3 Conclusions

In summary, the work in this thesis demonstrates that the ambiguity disadvantage in word comprehension does not arise due to decision-making processes. Rather, it arises whenever there is competition between multiple semantic representations in the race for activation, as inherently predicted by the distributed view of the lexicon. It is important to note that this process is not entirely unique to the semantic level as competition for activation has also been observed for phonologically (e.g., Gaskell & Marslen-Wilson, 2002; Henderson, Weighall, Brown, & Gaskell, 2013; McClelland & Elman, 1986; Norris, 1994) and orthographically similar word forms (e.g., Bowers et al., 2005; Davis & Taft, 2005; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981). This leads to the conclusion that competition between lexical-semantic representations is the rule rather than the exception, owing to the way our knowledge of words and their meanings is stored and accessed. The present work reveals that semantic competition occurs both for familiar and newly-learnt meanings, provided that the meanings have separate semantic representations that are initially activated in parallel. This is true for homonyms, particularly balanced ones, and irregular polysemes, but not for metaphors and metonyms. For the latter two, it appears that the mental lexicon does not explicitly store alternative senses but only the rules that serve to generate these senses on-line. These rules reduce the need to store information in the lexicon and provide a productive mechanism that allows children and adults to work out the meaning of novel metaphorical and metonymic senses with little difficulty.

References

Ahrens, K., Chang, L. L., Chen, K. J., & Huang, C. R. (1998). Meaning representation and meaning instantiation for Chinese nominals. *Computational Linguistics & Chinese Language Processing, 3*(1), 45-60.

Amsel, B. D. (2011). Tracking real-time activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, *49*(5), 970-983. DOI: 10.1016/j.neuropsychologia.2011.01.003

Annett, M. (1967). The binomial distribution of right, mixed and left handedness. *The Quarterly Journal of Experimental Psychology*, *19*(4), 327-333. DOI: 10.1080/ 14640746708400109

Apresjan, J. D. (1974). Regular polysemy. *Linguistics, 12*(142), 5-32. DOI: 10.1515/ling.1974.12.142.5

Armstrong, B. C. (2016). *Number of meanings, number of senses, and part-of-speech frequencies from the Wordsmyth Online Dictionary* [data file]. Retrieved from https://cnbc.cmu.edu/~armstrong/tools/

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. (2011). Inducing homonymy effects via stimulus quality and (not) nonword difficulty: Implications for models of semantic ambiguity and word recognition. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Expanding the space of cognitive science: Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2223-2228). 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. DOI: 10.1080/23273798.2016.1171366

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. DOI: 10.3758/s13428-012-0199-8

Armstrong, B. C., Zugarramurdi, C., Cabana, Á., Lisboa, J. V., & Plaut, D. C. (2015). Relative meaning frequencies for 578 homonyms in two Spanish dialects: A cross-linguistic extension of the English eDom norms. *Behavior Research Methods, 48*, 1-13. DOI: 10.3758/s13428-015-0639-3

Asher, N. (2011). *Lexical meaning in context*. Cambridge, UK: Cambridge University Press.

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. DOI: 10.1016/S0278-2626(03)00095-2

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. DOI: 10.1006/jmla.1997.2502

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. DOI: 10.1016/j.jml.2007.12.005

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283-316. DOI: 10.1037/0096-3445.133.2.283

Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 827-845. DOI: 10.1037/0278-7393.22.4.827

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). In M. J. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd Edition, pp. 285-375). San Diego, CA: Academic Press.

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. DOI: 10.1016/j.jml.2012.11.001

Bartoń, K. (2014). Multi-model inference [R package]. Retrieved from http://cran.r-project.org/web/packages/MuMIn/index.html

Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology*, *52*(1), 369-396. DOI: 10.1146/ annurev.psych.52.1.369

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.

Bedny, M., McGill, M., & Thompson-Schill, S. L. (2008). Semantic adaptation and competition during word comprehension. *Cerebral Cortex*, *18*(11), 2574-2585. DOI: 10.1093/cercor/bhn018

Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, *30*(2), 161-169. DOI: 10.1111/j.1469-8986.1993.tb01729.x

Betts, H. N., Gilbert, R. A., Cai, Z. G., Okedara, Z. B., & Rodd, J. M. (2017). Retuning of lexical-semantic representations: Repetition and spacing effects in word-meaning priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(7), 1130-1150. DOI: 10.1037/ xlm0000507

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. DOI: 10.1016/j.cogbrainres.2004.12.006

Bilenko, N. Y., Grindrod, C. M., Myers, E. B., & Blumstein, S. E. (2008). Neural correlates of semantic competition during processing of ambiguous words. *Journal of Cognitive Neuroscience*, *21*(5), 960-975. DOI: 10.1162/ jocn.2009.21073

Binder, K. S., & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1186-1196. DOI: 10.1037/0278-7393.21.5.1186

Binder, K. S., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin & Review*, *5*(2), 271-276. DOI: 10.3758/BF03212950

Bitan, T., Kaftory, A., Meiri-Leib, A., Eviatar, Z., & Peleg, O. (2017). Phonological ambiguity modulates resolution of semantic ambiguity during reading: An fMRI study of Hebrew. *Neuropsychology*, *31*(7), 759-777. DOI: 10.1037/neu0000357

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353-1366. DOI: 10.1162/0898929054985473

Booth, J. R., Harasaki, Y., & Burman, D. D. (2006). Development of lexical and sentence level context effects for dominant and subordinate word meanings of homonyms. *Journal of Psycholinguistic Research*, *35*(6), 531-554. DOI: 10.1007/s10936-006-9028-5

Borowsky, R., & Masson, M. E. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 63-85.

Bott, L., Rees, A., & Frisson, S. (2016). The time course of familiar metonymy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(7), 1160-1170. DOI: 10.1037/xlm0000218

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*(3), B45-B54. DOI: 10.1016/j.cognition.2005.02.002

Briggs, G. G., & Nebes, R. D. (1975). Patterns of hand preference in a student population. *Cortex*, *11*(3), 230-238. DOI: 10.1016/S0010-9452(75)80005

British National Corpus (2007). *British National Corpus*. Oxford, UK: Oxford University Computing Services.

Brocher, A., Foraker, S., & Koenig, J. P. (2016). Processing of irregular polysemes in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1798-1813. DOI: 10.1037/xlm0000271

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. DOI: 10.1080/23273798.2017.1381748

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

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34-44. DOI: 10.1162/jocn.1993.5.1.34

Burgess, C., & Simpson, G. B. (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language*, *33*(1), 86-103. DOI: 10.1016/0093-934X(88)90056-9

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, *98*, 73-101. DOI: 10.1016/ j.cogpsych.2017.08.003

Casenhiser, D. M. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language*, *32*(2), 319-343. DOI: 10.1017/S0305000904006749

Caramazza, A., & Grober, E. (1976). Polysemy and the structure of the subjective lexicon. In C. Rameh (Ed.), *Georgetown University roundtable on languages and linguistics. Semantics: Theory and application* (pp. 181-206). Washington, DC: Georgetown University Press.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5-51. DOI: 10.1177/1529100618772271

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. DOI: 10.1016/j.jneumeth.2015.02.025

Chen, L., & Boland, J. E. (2008). Dominance and context effects on activation of alternative homophone meanings. *Memory & Cognition*, *36*(7), 1306-1323. DOI: 10.3758/MC.36.7.1306

Chwilla, D. J., & Kolk, H. H. (2003). Event-related potential and reaction time evidence for inhibition between alternative meanings of ambiguous words. *Brain and Language*, *86*(2), 167-192. DOI: 10.1016/S0093-934X(02)00527-8

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 55(4), 767-811. DOI: 10.2307/412745

Clark, H. H., & Gerrig, R. J. (1983). Understanding old words with new meanings. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 591-608. DOI: 10.1016/S0022-5371(83)90364-X

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428. DOI: 10.1037/0033-295X.82.6.407

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497-505. DOI: 10.1080/14640748108400805

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, *4*(1), 5-17. DOI: 10.1080/13603110600574322

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204-256. DOI: 10.1037/0033-295X.108.1.204

Copestake, A., & Briscoe, T. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics, 12*(1), 15-67. DOI: 10.1093/jos/12.1.15

Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: Homophony and the distribution of learning exemplars. *Language Learning and Development*, *12*(3), 231-251. DOI: 10.1080/15475441.2015.1127163

Davenport, T., & Coulson, S. (2011). Predictability and novelty in literal language comprehension: An ERP study. *Brain Research*, *1418*, 70-82. DOI: 10.1016/ j.brainres.2011.07.039

Davies, C., Porretta, V., Koleva, K., & Klepousniotou, E. (2017, September). *Do speaker-specific cues influence ambiguous word interpretation?* Poster presented at the 23rd Architectures and Mechanisms of Language Processing Conference, Lancaster, UK.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*(1), 65-70. DOI: 10.3758/BF03206399

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3773-3800. DOI: 10.1098/rstb.2009.0111

Davis, C. J., & Taft, M. (2005). More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin & Review*, *12*(5), 904-910. DOI: 10.3758/BF03196784

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. DOI: 10.1016/ j.jneumeth.2003.10.009

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. DOI: 10.3389/fpsyg.2015.00077

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. DOI: 10.1111/psyp.12762

Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*(1), 203-214. DOI: 10.1017/S030500090300583X

Duffy, S. A., Kambe, G., & Rayner, K. (2001). The effect of prior disambiguating context on the comprehension of ambiguous words: Evidence from eye movements. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 27-43). Washington, DC: American Psychological Association.

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. DOI: 10.1016/ 0749-596X(88)90066-6

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. DOI: 10.3758/ s13423-014-0665-7

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301-306. DOI: 10.1016/j.tics.2004.05.003

Elston-Güttler, K. E., & Friederici, A. D. (2007). Ambiguous words in sentences: Brain indices for native and non-native disambiguation. *Neuroscience Letters*, *414*(1), 85-89. DOI: 10.1016/j.neulet.2006.12.002

Falkum, I. L., & Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua, 157*, 1-16.

Fang, X., & Perfetti, C. A. (2017). Perturbation of old knowledge precedes integration of new knowledge. *Neuropsychologia*, *99*, 270-278. DOI: 10.1016/j.neuropsychologia.2017.03.015

Fang, X., & Perfetti, C. A. (2018). Learning new meanings for known words: Perturbation of original meanings and retention of new meanings. *Memory & cognition*, 1-15. DOI: 10.3758/s13421-018-0855-z

Fang, X., Perfetti, C., & Stafura, J. (2017). Learning new meanings for known words: Biphasic effects of prior knowledge. *Language, Cognition and Neuroscience*, *32*(5), 637-649. DOI: 10.1080/23273798.2016.1252050

Faust, M., & Lavidor, M. (2003). Semantically convergent and semantically divergent priming in the cerebral hemispheres: Lexical decision and semantic judgment. *Cognitive Brain Research*, *17*(3), 585-597. DOI: 10.1016/S0926-6410(03)00172-1

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.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Fiebach, C. J., Vos, S. H., & Friederici, A. D. (2004). Neural correlates of syntactic ambiguity in sentence comprehension for low and high span readers. *Journal of Cognitive Neuroscience*, *16*(9), 1562-1575. DOI: 10.1162/0898929042568479

Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, *116*(3), 421-436. DOI: 10.1016/j.cognition.2010.06.005

Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language*, *67*(4), 407-425. DOI: 10.1016/j.jml.2012.07.010

Forster, K. I. (2004). Category size effects revisited: Frequency and masked priming effects in semantic categorization. *Brain and Language*, *90*(1), 276-286. DOI: 10.1016/S0093-934X(03)00440-1

Forster, K. I., & Hector, J. (2002). Cascaded versus noncascaded models of lexical and semantic processing: The turple effect. *Memory & Cognition*, *30*(7), 1106-1117. DOI: 10.3758/BF03194328

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*(2), 181-200. DOI: 10.1016/0749-596X(90)90071-7

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research*, *11*(2), 305-323. DOI: 10.1016/S0926-6410(00)00065-3

Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass*, *3*(1), 111-127. DOI: 10.1111/j.1749-818X.2008.00104.x

Frisson, S., & Frazier, L. (2005). Carving up word meaning: Portioning and grinding. *Journal of Memory and Language*, *53*(2), 277-291. DOI: 10.1016/j.jml.2005.03.004

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1366-1383. DOI: 10.1037/0278-7393.25.6.1366

Frisson, S., & Pickering, M. J. (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes*, *22*(4), 595-613. DOI: 10.1080/01690960601017013

Frömer, R., Maier, M., & Abdel Rahman, R. (2018). Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. *Frontiers in Neuroscience*, *12*, 48. DOI: 10.3389/fnins.2018.00048

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.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105-132. DOI: 10.1016/S0010-0277(03)00070-2

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*(5/6), 613-656. DOI: 10.1080/016909697386646

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220-266. DOI: 10.1016/S0010-0285(02)00003-8

Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition*, *182*, 109-126*. DOI:* 10.1016/j.cognition.2018.09.007

Geeraerts, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics*, *4*(3), 223-272. DOI: 10.1515/cogl.1993.4.3.223

Geis, M. F., & Winograd, E. (1974). Norms of semantic encoding variability for fifty homographs. *Bulletin of the Psychonomic Society*, *3*(6), 429-431. DOI: 10.3758/BF03333516

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256-281. DOI: 10.1037/0096-3445.113.2.256

Gilbert, R. A., Betts, H. N., Jose, R., & Rodd, J. M. (2017, January). New UK-based dominance norms for spoken ambiguous words. Poster presented at the Annual Experimental Psychology Society Conference, London, UK.

Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2018). Listeners and readers generalize their experience with word meanings across modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(10), 1533-1561. DOI: 10.1037/xlm0000532

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183-206. DOI: 10.1515/ cogl.1997.8.3.183

Glucksberg, S., Kreuz, R. J., & Rho, S. H. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 323-335. DOI: 10.1037/ 0278-7393.12.3.323

Goddard, C. (1998). *Semantic analysis: A practical introduction*. Oxford, UK: Oxford University Press.

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. DOI: 10.1037/0096-1523.25.2.561

Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1158-1176. DOI: 10.1037/0096-1523.20.6.1158

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*(3), 518-565. DOI: 10.1037/0033-295X.103.3.518

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. DOI: 10.1016/ j.bandl.2014.10.001

Hagoort, P. (1993). Impairments of lexical-semantic processing in aphasia: Evidence from the processing of lexical ambiguities. *Brain and Language*, *45*(2), 189-232. DOI: 10.1006/brln.1993.1043

Hala, S., Pexman, P. M., & Glenwright, M. (2007). Priming the meaning of homographs in typically developing children and children with autism. *Journal of Autism and Developmental Disorders*, *37*(2), 329-340. DOI: 10.1007/s10803-006-0162-6

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. DOI: 10.1027/ 1618-3169/a000062

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. DOI: 10.1017/S0142716416000266

Haro, J., & Ferré, P. (2018). Semantic ambiguity: Do multiple meanings inhibit or facilitate word recognition? *Journal of Psycholinguistic Research*, *47*(3), 679-698. DOI: 10.1007/s10936-017-9554-3

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491-528. DOI: 10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662-720. DOI: 10.1037/0033-295X.111.3.662

Harpaz, Y., Lavidor, M., & Goldstein, A. (2013). Right semantic modulation of early MEG components during ambiguity resolution. *Neuroimage*, *82*, 107-114. DOI: 10.1016/j.neuroimage.2013.05.060

Harper, D. (2016). Online Etymology Dictionary. Retrieved from http:// etymonline.com

Henderson, L., Snowling, M., & Clarke, P. (2013). Accessing, integrating, and inhibiting word meaning in poor comprehenders. *Scientific Studies of Reading*, *17*(3), 177-198. DOI: 10.1080/10888438.2011.652721

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online lexical competition during spoken word recognition and word learning in children and adults. *Child Development*, *84*(5), 1668-1685. DOI: 10.1111/ cdev.12067

Hill, H., Strube, M., Roesch-Ely, D., & Weisbrod, M. (2002). Automatic vs. controlled processes in semantic priming - Differentiation by event-related potentials. *International Journal of Psychophysiology*, *44*(3), 197-218. DOI: 10.1016/S0167-8760(01)00202-1

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental*

Psychology: Human Perception and Performance, *22*(6), 13311-1356. DOI: 10.1037/0096-1523.22.6.1331

Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 686-713. DOI: 10.1037/0278-7393.28.4.686

Hino, Y., Kusunose, Y., & Lupker, S. J. (2010). The relatedness-of-meaning effect for ambiguous words in lexical-decision tasks: When does relatedness matter? *Canadian Journal of Experimental Psychology, 64*(3), 180-196. DOI: 10.1037/a0020475

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. DOI: 10.1016/j.jml.2006.04.001

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*(1), 74-95. DOI: 10.1037/0033-295X.98.1.74

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., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718-730. DOI: 10.3758/s13428-012-0278-x

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. DOI: 10.1037/a0038995

Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *14*(3), 265-274. DOI: 10.1016/S0022-5371(75)80070-3

Huettig, F., & Altmann, G. T. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 985-1018. DOI: 10.1080/13506280601130875

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114-126. DOI: 10.1016/j.jml.2018.08.005

Ingram, J., Hand, C. J., & Maciejewski, G. (2016). Exploring the measurement of markedness and its relationship with other linguistic variables. *PloS One*, *11*(6), e0157141. DOI: 10.1371/journal.pone.0157141

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431-439. DOI: 10.3758/BF03208203

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. DOI: 10.1075/ml.10.1.05jag

Jager, B., & Cleland, A. A. (2016). Polysemy advantage with abstract but not concrete words. *Journal of Psycholinguistic Research*, *45*(1), 143-156. DOI: 10.1007/s10936-014-9337-z

Jager, B., Green, M. J., & Cleland, A. A. (2016). Polysemy in the mental lexicon: Relatedness and frequency affect representational overlap. *Language, Cognition and Neuroscience*, *31*(3), 425-429. DOI: 10.1080/23273798.2015.1105986

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology*, *13*(2), 278-305. DOI: 10.1016/0010-0285(81)90011-6

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944-946. DOI: 10.1111/2041-210X.12225

Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1051-1062. DOI: 10.1037/0278-7393.20.5.1051

Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research*, *23*(6), 459-471. DOI: 10.1007/BF02146685

Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory & Cognition*, *29*(2), 363-372. DOI: 10.3758/ BF03194931

Kandhadai, P., & Federmeier, K. D. (2007). Multiple priming of lexically ambiguous and unambiguous targets in the cerebral hemispheres: The coarse coding hypothesis revisited. *Brain Research*, *1153*, 144-157. DOI: 10.1016/j.brainres.2007.03.046

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. DOI: 10.1006/jmla.1993.1026

Kawamoto, A. H., Farrar, W. T., & Kello, C. T. (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1233-1247. DOI: 10.1037/0096-1523.20.6.1233

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *The Quarterly Journal of Experimental Psychology*, *63*(1), 160-193. DOI: 10.1080/17470210902866664

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*(2), 259-282. DOI: 10.1006/ jmla.2001.2779

Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations of polysemous senses. *Journal of Memory and Language*, *47*(4), 548-570. DOI: 10.1016/S0749-596X(02)00020-7

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*(1), 205-223. DOI: 10.1006/brln.2001.2518

Klepousniotou, E., & Baum, S. R. (2005a). Unilateral brain damage effects on processing homonymous and polysemous words. *Brain and Language*, *93*(3), 308-326. DOI: 10.1016/j.bandl.2004.10.011

Klepousniotou, E., & Baum, S. R. (2005b). Processing homonymy and polysemy: Effects of sentential context and time-course following unilateral brain damage. *Brain and Language*, *95*(3), 365-382. DOI: 10.1016/ j.bandl.2005.03.001

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. DOI: 10.1016/j.jneuroling.2006.02.001

Klepousniotou, E., Gracco, V., & Pike, G. B. (2007, May). Competition effects in lexical ambiguity processing. Poster presented at the 14th Cognitive Neuroscience Society Conference, New York, USA.

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. DOI: 10.1016/ j.bandl.2013.06.002

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. DOI: 10.1016/ j.bandl.2012.06.007

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. DOI: 10.1037/a0013012

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. DOI: 10.1162/jocn_a_00977

Kostić, N. (2015). Antonym sequence in written discourse: A corpus-based study. *Language Sciences*, *47*, 18-31. DOI: 10.1016/ j.langsci.2014.07.013

Kotchoubey, B., & El-Khoury, S. (2014). Event-related potentials indicate context effect in reading ambiguous words. *Brain and Cognition*, *92*, 48-60. DOI: 10.1016/j.bandc.2014.10.007

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978-990. DOI: 10.3758/s13428-012-0210-4

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. DOI: 10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language*, *11*(2), 354-373. DOI: 10.1016/0093-934X(80)90133-9

Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, *1*(1), 38-49. DOI: 10.1162/jocn.1989.1.1.38

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: *Event-related brain potential investigations*. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83-143). San Diego, CA: Academic Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. DOI: 10.18637/jss.v082.i13

Lakoff, G. (1987). *Women, fire, and dangerous Things: What categories reveal about the mind*. Chicago, IL: The University of Chicago Press.

Landauer, T. K., & Freedman, J. L. (1968). Information retrieval from long-term memory: Category size and recognition time. *Journal of Verbal Learning and Verbal Behavior*, *7*(2), 291-295. DOI: 10.1016/S0022-5371(68)80003

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, *48*(2), 176-186. DOI: 10.1111/j.1469-8986.2010.01058.x

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306-353. DOI: 10.1016/j.cogpsych.2007.01.001

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. DOI: 10.1016/j.jml.2016.09.002

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. DOI: 10.1016/j.brainres.2006.01.058

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. DOI: 10.1016/j.jml.2009.08.003

Lee, C. L., & Federmeier, K. D. (2011). Differential age effects on lexical ambiguity resolution mechanisms. *Psychophysiology*, *48*(7), 960-972. DOI: 10.1111/j.1469-8986.2010.01158.x

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics*, *1*(2), 207-246. DOI: 10.1515/ cogl.1990.1.2.207

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38. DOI: 10.1017/S0140525X99001776

Li, L., & Slevc, L. R. (2017). Of papers and pens: Polysemes and homophones in lexical (mis) selection. *Cognitive Science*, *41*(S6), 1532-1548. DOI: 10.1111/ cogs.12402

Lin, C. C., & Ahrens, K. (2005). How many meanings does a word have? Meaning estimation in Chinese and English. In J. Menitt, & W. S-Y. Wang (Eds.), *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics* (pp. 437-464). Hong Kong: City University of Hong Kong Press.

Lin, C. J. C., & Ahrens, K. (2010). Ambiguity advantage revisited: Two meanings are better than one when accessing Chinese nouns. *Journal of Psycholinguistic Research*, *39*(1), 1-19. DOI: 10.1007/s10936-009-9120

Locker, L., Simpson, G. B., & Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. *Memory & Cognition*, *31*(4), 505-515. DOI: 10.3758/BF03196092

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476-490. DOI: 10.3758/BF03210951

Lopukhina, A., Laurinavichyute, A., Lopukhin, K., & Dragoy, O. (2018). The mental representation of polysemy across word classes. *Frontiers in Psychology*, *9*, 192. DOI: 10.3389/fpsyg.2018.00192

Lowder, M. W., & Gordon, P. C. (2013). It's hard to offend the college: Effects of sentence structure on figurative-language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 993-1011. DOI: 10.1037/a0031671

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618-630. DOI: 10.3758/ BF03212999

Lupker, S. J. (2007). Representation and processing of lexically ambiguous words. In M. G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp.159-174). Oxford, UK: Oxford University Press.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge, UK: Cambridge University Press.

MacDonald, M. C. (1997). Lexical representations and sentence processing: An introduction. *Language and Cognitive Processes*, *12*(2-3), 121-136. DOI: 10.1080/016909697386826

MacGregor, L. J., Bouwsema, J., & Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. *Neuropsychologia*, *68*,126-138. DOI: 10.1016/ j.neuropsychologia.2015

Maciejewski, G. (2014). *Does meaning acquisition entail meaning competition?* (Unpublished dissertation). University College London, London, UK.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 148-172). Cambridge, MA: MIT Press.

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2008). Hemispheric differences in processing the literal interpretation of idioms: Converging evidence from behavioral and fMRI studies. *Cortex*, *44*(7), 848-860. DOI: 10.1016/j.cortex.2007.04.004

Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research*, *1146*, 115-127. DOI: 10.1016/ j.brainres.2007

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. DOI: 10.1016/j.jml.2017.01.001

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86. DOI: 10.1016/0010-0285(86)90015-0

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375-407. DOI: 10.1037/0033-295X.88.5.375

McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science*, *30*(1), 181-192. DOI: 10.1207/s15516709cog0000_49

Meade, G., & Coch, D. (2017). Word-pair priming with biased homonyms: N400 and LPC effects. *Journal of Neurolinguistics*, *41*, 24-37. DOI: 10.1016/ j.jneuroling.2016.09.002

Medler, D. A., & Binder, J. R. (2005). MCWord: An on-line orthographic database of the English language. Retrieved from http:// neuro.mcw.edu/mcword

Millis, M. L., & Bution, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don’t. *Memory & Cognition*, *17*(2), 141-147. DOI: 10.3758/BF03197064

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. DOI: 10.1111/j.1551-6709.2009.01069.x

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. DOI: 10.1016/ j.bandl.2018.01.001

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165-178. DOI: 10.1037/h0027366

Murphy, G. L. (1997). Polysemy and the creation of novel word meanings. In T. B. Ward, S. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 235–265). Washington, DC: American Psychological Association.

Murphy, G. L. (2006). Comprehending new words beyond their original contexts. *Skase Journal of Theoretical Linguistics*, *3*(2), 2-8.

Musz, E., & Thompson-Schill, S. L. (2017). Tracking competition and cognitive control during language comprehension with multi-voxel pattern analysis. *Brain and Language*, *165*, 21-32. DOI: 10.1016/ j.bandl.2016.11.002

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142. DOI: 10.1111/j.2041-210x.2012.00261.x

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(1), 1-17. DOI: 10.3758/CABN.7.1.1

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226-254. DOI: 10.1037/0096-3445.106.3.226

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. DOI: 10.3758/ BF03195588

Nelson, D. L., McEvoy, C. L., Walling, J. R., & Wheeler, J. W. (1980). The University of South Florida homograph norms. *Behavior Research Methods & Instrumentation*, *12*(1), 16-37. DOI: 10.3758/BF03208320

Nieuwland, M. S., & van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098-1111. DOI: 10.1162/jocn.2006.18.7.1098

Nievas, F., & Justicia, F. (2004). A cross-sectional study about meaning access processes for homographs. *Cognitive Development*, *19*(1), 95-109. DOI: 10.1016/j.cogdev.2003.11.002

Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology*, *90*(2), 142-171. DOI: 10.1016/j.jecp.2004.11.003

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234. DOI: 10.1016/0010-0277(94)90043-4

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. DOI: 10.1080/ 02643290903519367

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263-281. DOI: 10.3758/CABN.5.3.263

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, *3*(2), 143-184. DOI: 10.1007/BF00126509

Oden, G. C., & Spira, J. L. (1983). Influence of context on the activation and selection of ambiguous word senses. *The Quarterly Journal of Experimental Psychology*, *35*(1), 51-64. DOI: 10.1080/14640748308402116

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, *9*(3), 225-236. DOI: 10.3758/ BF03196957

Parks, R., Ray, J., & Bland, S. (1998). Wordsmyth English dictionary-thesaurus. Chicago, IL: The University of Chicago Press. Retrieved from http:// wordsmyth.net

Pecher, D., Zeelenberg, R., & Wagenmakers, E. J. (2005). Enemies and friends in the neighborhood: Orthographic similarity effects in semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 121-128. DOI: 10.1037/0278-7393.31.1.121

Peleg, O., Giora, R., & Fein, O. (2008). Resisting contextual information: You can't put a salient meaning down. *Lodz Papers in Pragmatics*, *4*(1), 13-44. DOI: 10.2478/ v10016-008-0008-0

Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition: Meaning and context, individuals and development* (pp. 24-43). Hove, UK: Psychology Press.

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. DOI: 10.1037/0278-7393.30.6.1252

Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 556-573. DOI: 10.1037/0278-7393.27.2.556

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. DOI: 10.3758/BF03201255

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects Models in S and S-PLUS*. New York, NY: Springer.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12*(5-6), 765-806. DOI: 10.1080/016909697386682

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56-115. DOI: 10.1037/ 0033-295X.103.1.56

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. DOI: 10.1037/0278-7393.21.3.785

Pustejovsky, J. (1995). *The generative lexicon.* Cambridge, MA: MIT Press.

Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, *18*(1), 97-109. DOI: 10.1162/089892906775250003

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, *7*(5), 187-189. DOI: 10.1016/S1364-6613(03)00092-5

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

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2011). Rules, radical pragmatics and restrictions on regular polysemy. *Journal of Semantics*, *28*(4), 485-512. DOI: 10.1093/jos/ffr005

Rabagliati, H., Pylkkänen, L., & Marcus, G. F. (2013). Top-down influence in young children's linguistic ambiguity resolution. *Developmental Psychology*, *49*(6), 1076-1089. DOI: 10.1037/a0026918

Rabagliati, H., & Snedeker, J. (2013). The truth about chickens and bats: Ambiguity avoidance distinguishes types of polysemy. *Psychological Science*, *24*(7), 1354-1360. DOI: 10.1177/0956797612472205

Rastle, K. (2007). Visual word recognition. In M. G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 71-87). Oxford, UK: Oxford University Press.

Rayner, K., Cook, A. E., Juhasz, B. J., & Frazier, L. (2006). Immediate disambiguation of lexically ambiguous words during reading: Evidence from eye movements. *British Journal of Psychology*, *97*(4), 467-482. DOI: 10.1348/ 000712605X89363

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. DOI: 10.3758/ BF03197692

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 779-790. DOI: 10.1037/0278-7393.15.5.779

Rice, C. A., Beekhuizen, B., Dubrovsky, V., Stevenson, S., & Armstrong, B. C. (2018). A comparison of homonym meaning frequency estimates derived from movie and television subtitles, free association, and explicit ratings. *Behavior Research Methods*, 1-27. DOI: 10.3758/s13428-018-1107-7

Rodd, J. M. (2004). When do leotards get their spots? Semantic activation of lexical neighbors in visual word recognition. *Psychonomic Bulletin & Review*, *11*(3), 434-439. DOI: 10.3758/BF03196591

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., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition*, *40*(7), 1095-1108. DOI: 10.3758/s13421-012-0209-1

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, *87*, 16-37. DOI: 10.1016/j.jml.2015.10.006

Rodd, J. M., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, *68*(2), 180-198. DOI: 10.1016/j.jml.2012.08.002

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*(8), 1261-1269. DOI: 10.1093/cercor/bhi009

Rodd, J., 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. DOI: doi:10.1006/jmla.2001.2810

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. DOI: 10.1207/s15516709cog2801_4

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2010). The role of domain-general frontal systems in language comprehension: Evidence from dual-task interference and semantic ambiguity. *Brain and Language*, *115*(3), 182-188. DOI: 10.1016/j.bandl.2010.07.005

Rodd, J. M., Longe, O. A., Randall, B., & Tyler, L. K. (2010). The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia*, *48*(5), 1324-1335. DOI: 10.1016/j.neuropsychologia.2009.12.035

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An activation likelihood estimation meta-analysis. *Brain and Language*, *141*, 89-102. DOI: 10.1016/j.bandl.2014.11.012

Rossell, S. L., Bullmore, E. T., Williams, S. C., & David, A. S. (2001). Brain activation during automatic and controlled processing of semantic relations: Priming using lexical decision. *Neuropsychologia*, *39*(11), 1167-1176. DOI: 10.1016/S0028-3932(01)00049-5

Rossell, S. L., Price, C. J., & Nobre, A. C. (2003). The anatomy and time course of semantic priming investigated by fMRI and ERPs. *Neuropsychologia*, *41*(5), 550-564. DOI: 10.1016/S0028-3932(02)00181-1

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 487-494. DOI: 10.1016/S0022-5371(70)80091-3

Ruhl, C. (1989). *On monosemy: A study in linguistic semantics*. Albany, NY: State University of New York Press.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 45-76). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60-94. DOI: 10.1037/0033-295X.89.1.60

Salisbury, D. (2010). N400 to lexical ambiguity and semantic incongruity in schizophrenia. *International Journal of Psychophysiology*, *75*(2), 127-132. DOI: 10.1016/j.ijpsycho.2009.10.002

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

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.

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.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568. DOI: 10.1037/0033-295X.96.4.523

Sereno, S. C., O'Donnell, P. J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and Performance, 32*(2), 335-350. DOI: 10.1037/0096-1523.32.2.335

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489-493. DOI: 10.1016/j.tics.2003.09.010

Siakaluk, P. D., Pexman, P. M., Sears, C. R., & Owen, W. J. (2007). Multiple meanings are not necessarily a disadvantage in semantic processing: Evidence from homophone effects in semantic categorisation. *Language and Cognitive Processes*, *22*(3), 453-467. DOI: 10.1080/01690960600834756

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 359-374). San Diego, CA: Academic Press.

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. DOI: 10.1037/0096-1523.11.1.28

Simpson, G. B., & Foster, M. R. (1986). Lexical ambiguity and children's word recognition. *Developmental Psychology*, *22*(2), 147-154. DOI: 10.1037/0012-1649.22.2.147

Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, *30*(6), 627-643. DOI: 10.1016/0749-596X(91)90029-J

Simpson, J., & Weiner, E. S. (1989). *Oxford English dictionary*. Oxford, UK: Clarendon Press. Retrieved from: http://en.oxforddictionaries.com

Srinivasan, M., Al-Mughairy, S., Foushee, R., & Barner, D. (2017). Learning language from within: Children use semantic generalizations to infer word meanings. *Cognition*, *159*, 11-24. DOI: 10.1016/j.cognition.2016.10.019

Srinivasan, M., Berner, C., & Rabagliati, H. (2018). Children’s use of polysemy to structure new word meanings. *Journal of Experimental Psychology: General*. DOI: 10.1037/xge0000454

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, *157*, 124-152. DOI: 10.1016/j.lingua.2014.12.004

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology*, *62*(4), 245-272. DOI: 10.1016/j.cogpsych.2011.03.002

Srinivasan, M., & Snedeker, J. (2014). Polysemy and the taxonomic constraint: Children's representation of words that label multiple. *Language Learning and Development, 10*(2), 1-32. DOI: 10.1080/15475441.2013.820121

Standards & Testing Agency (2016). Key Stage 2 English Reading Booklet. Retrieved from http://gov.uk/govenrment/publications/key-stage-2-tests-2016-english-reading-test-materials/

Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*, *32*(4), 827-853. DOI: 10.1017/ S0305000905007099

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. DOI: 10.1016/S0093-934X(02)00547-3

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language*, *27*(3), 324-340. DOI: 10.1016/ 0749-596X(88)90058-7

Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language*, *32*(3), 359-372. DOI: 10.1006/jmla.1993.1019

Taler, V., Kousaie, S., & López Zunini, R. (2013). ERP measures of semantic richness: The case of multiple senses. *Frontiers in Human Neuroscience*, *7*, 5. DOI: 10.3389/fnhum.2013.00005

Tamminen, J., Cleland, A. A., Quinlan, P. T., & Gaskell, M. G. (2006). Processing semantic ambiguity: Different loci for meanings and senses. In R. Sun, & N. Miyaki (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2222-2227). Mahwah, NJ: Lawrence Erlbaum Associates.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, *30*(43), 14356-14360. DOI: 10.1523/JNEUROSCI.3028-10.2010

Taylor, J. S. H., Duff, F. J., Woollams, A. M., Monaghan, P., & Ricketts, J. (2015). How word meaning influences word reading. *Current Directions in Psychological Science*, *24*(4), 322-328. DOI: 10.1177/ 0963721415574980

Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language*, *38*(4), 440-458. DOI: 10.1006/jmla.1997.2559

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 reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792-14797. DOI: 10.1073/pnas.94.26.14792

Titone, D., Levy, D. L., & Holzman, P. S. (2000). Contextual insensitivity in schizophrenic language processing: Evidence from lexical ambiguity. *Journal of Abnormal Psychology*, *109*(4), 761-767. DOI: 10.1037/0021-843X.109.4.761

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes*, *22*(5), 727-779. DOI: 10.1080/01690960601057068

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*(4), 566-585. DOI: 10.1006/jmla.1996.0030

Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychonomic Bulletin & Review, 7*(1), 49-82. DOI: 10.3758/BF03210725

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. DOI: 10.3758/BF03202766

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657-671. DOI: 10.1162/089892999563724

van der Schoot, M., Vasbinder, A. L., Horsley, T. M., Reijntjes, A., & van Lieshout, E. C. (2009). Lexical ambiguity resolution in good and poor comprehenders: An eye fixation and self-paced reading study in primary school children. *Journal of Educational Psychology*, *101*(1), 21-36. DOI: 10.1037/a0013382

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, *97*(4), 488-522.

Vitello, S., & Rodd, J. M. (2015). Resolving semantic ambiguities in sentences: Cognitive processes and brain mechanisms. *Language and Linguistics Compass*, *9*(10), 391-405. DOI: 10.1111/lnc3.12160

Vitello, S., Warren, J. E., Devlin, J. T., & Rodd, J. M. (2014). Roles of frontal and temporal regions in reinterpreting ambiguous sentences. *Frontiers in Human Neuroscience*, *8*, 530. DOI: 10.3389/ fnhum.2014.00530

Vu, H., Kellas, G., Metcalf, K., & Herman, R. (2000). The influence of global discourse on lexical ambiguity resolution. *Memory & Cognition*, *28*(2), 236-252. DOI: 10.3758/BF03213803

Vu, H., Kellas, G., & Paul, S. T. (1998). Sources of sentence constraint on lexical ambiguity resolution. *Memory & Cognition*, *26*(5), 979-1001. DOI: 10.3758 /BF03201178

Weinreich, U. (1966). Explorations in semantic theory. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 3, pp. 395-477). The Hague, the Netherlands: Mouton.

Whitney, C., Jefferies, E., & Kircher, T. (2011). Heterogeneity of the left temporal lobe in semantic representation and control: Priming multiple versus single meanings of ambiguous words. *Cerebral Cortex*, *21*(4), 831-844. DOI: 10.1093/cercor/bhq148

Williams, J. N. (1992). Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research*, *21*(3), 193-218. DOI: 10.1007/BF01068072

Witzel, J., & Forster, K. (2014). Lexical co-occurrence and ambiguity resolution. *Language, Cognition and Neuroscience*, *29*(2), 158-185. DOI: 10.1080/01690965.2012.748925

Woollams, A. M. (2015). Lexical is as lexical does: Computational approaches to lexical representation. *Language, Cognition and Neuroscience*, *30*(4), 395-408. DOI: 10.1080/23273798.2015.1005637

Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: Eye movements are sensitive to overlapping semantic features, not association. *Psychonomic Bulletin & Review*, *16*(5), 869-874. DOI: 10.3758/PBR.16.5.869

Zempleni, M. Z., Renken, R., Hoeks, J. C., Hoogduin, J. M., & Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *Neuroimage*, *34*(3), 1270-1279. DOI: 10.1016/ j.neuroimage.2006

Appendices

Appendix 1: Experiment 1: 100 homonyms included in the British-English norms of relative meaning frequency. "Meaning$_1$" and "Meaning$_2$" refer to the first and the second word meaning listed in the Wordsmyth Dictionary (Parks et al., 1998). "Frequency" represents means of meaning-frequency ratings provided by 100 participants. Greater scores of meaning dominance (β) denote more unbalanced homonymy. Meaning relatedness was rated on a scale from 1 ("highly unrelated") to 7 ("highly related").

Word	Meaning$_1$		Meaning$_2$		Meaning dominance (β)	Meaning relatedness
	Definition	Frequency (%)	Definition	Frequency (%)		
angle	the geometric figure made by two lines extending out from a single point	79.71	to fish with hook and line	19.25	0.72	2.03
arch	a structural element, usually of masonry, that is curved and used to span an open space such as a door	90.58	mischievous or sly	9.12	0.89	1.83
bay	a body of water partly surrounded by land, especially a recessed area along a shoreline; large cove	63.18	an interior recess in a wall, often containing a window	33.32	0.50	2.93
bear	to carry; to endure	30.72	a large, usually omnivorous, furry mammal with a short tail	68.13	0.55	1.57
blow	a quick, forcible hit by the fist or by a hard object	45.75	to be in swift motion, as the air or wind	52.82	0.38	3.97
buffer	a device, such as a bumper, that absorbs the force of a collision	38.47	something used to polish or shine	53.55	0.45	2.57
bush	a low plant having many woody branches; shrub	92.80	a lining that is used in machines to reduce friction	7.15	0.91	2.03
bust	a sculpture or other representation of the upper portion of the human body	55.08	to break or burst	44.93	0.51	1.87
calf	the young of cattle or of other bovine mammals, and of some other large mammals	47.57	the muscular back part of a human's leg below the knee	52.43	0.32	2.07

camp	an outdoor place where tents or temporary shelters are set up	66.32	the style of something such as clothing, decoration, art, or the like that is considered amusing because of its consciously pretentious showiness or outlandishness	32.93	0.49	1.77
cape	a sleeveless, loose-fitting garment that fastens at the neck and hangs loosely from the shoulders	77.02	a point of land jutting into a large body of water	22.98	0.72	1.73
card	a small piece of thick paper, cardboard, or plastic printed with personal data such as name and address and used as identification	93.84	a mechanical device used for combing cotton, wool, or the like in order to remove the shortest fibres prior to spinning	6.05	0.93	2.27
chord	a line segment connecting two points on a curve	16.68	three or more musical tones played at the same time	82.62	0.79	2.10
clock	a mechanical or electric device, other than a watch, for measuring or indicating time	92.81	a small emblem embroidered or woven on the side of a sock or stocking	2.21	0.97	1.80
corn	a tall cereal plant that produces cylinder-shaped ears with rows of edible yellow or white seeds	79.91	a small area of hard, calloused skin on the toe or foot	20.09	0.70	1.53
crash	(of a vehicle or moving object) to strike violently (against or through something)	90.69	a coarsely woven fabric made of irregular or rough yarn	5.52	0.95	1.60
dam	a barrier built usually across a waterway to restrict flow and raise the water level	91.69	a female parent, especially of a four-legged mammal	5.82	0.93	1.47
ear	the organ of hearing in man and vertebrate animals	85.78	the seed-bearing part of a plant such as corn	14.12	0.87	1.93
egg	ovum	78.86	to incite or encourage to act	21.14	0.70	1.60
fan	a hand-held device that opens out to form a triangular shape	45.87	an enthusiastic follower of an activity such as a sport of a	54.13	0.21	1.63

	and that is used to cool the face by waving back and forth		person or persons who engage in that activity			
fleet	a group of naval ships under one command or grouped for one purpose	75.74	swift and nimble	24.26	0.68	2.63
flight	an act of passing through air or space by flying	72.87	an act or instance of fleeing	27.08	0.59	4.63
flock	a group of animals or birds of one kind keeping or kept together, such as geese or sheep	87.53	a tuft, as of wool or other fibre	12.38	0.86	2.47
fly	to move through the air by means of wings	54.75	any of a variety of small, winged insects, especially the common housefly	42.67	0.21	5.53
forge	a furnace or hearth where metal is heated to be worked or shaped	43.36	to move ahead gradually but with determination	52.45	0.52	2.13
fry	to cook in hot butter, oil, or other fat	90.23	very young fish	9.76	0.87	2.13
gin	an alcoholic liquor made from grain distilled with juniper berries	92.23	a machine designed to remove the seeds from raw cotton; cotton gin	7.47	0.91	1.67
hide	to put or hold out of sight; keep from view; conceal	75.98	the skin of one of the larger animals such as a buffalo or cow	23.92	0.66	1.63
hiding	the act or condition of concealing or being concealed	74.38	a thrashing, flogging, or severe beating	25.37	0.65	1.83
host	a person who provides hospitality such as food, entertainment, or lodging for guests	81.69	a very large number of people or things	17.99	0.76	3.33
jam	to force or pack tightly into a small space	38.01	a sweet spread usually made by cooking fruit and sugar	58.71	0.39	2.23
keen	extremely sharp; able to cut	44.09	a mournful wailing for the dead	7.74	0.78	1.77

launch	to propel with force	87.06	a large open boat propelled by a motor	12.94	0.83	4.37
lawn	a stretch or plot of ground planted with grass or other low-growing ground covers and usually mowed short	90.35	a fine, sheer, woven fabric, usually of cotton or linen	9.65	0.92	2.27
lean	to bend or incline	59.18	having little flesh	40.62	0.39	1.70
limp	to walk in an uneven, laboured, lame manner	53.88	not stiff, rigid, or firm; floppy or flabby	46.12	0.32	4.07
lock	a mechanical device for preventing entry through a door or window or into a safe, usually opened with a key or combination	81.86	a curl or portion of hair	18.14	0.75	1.43
mail	the system organized to send and deliver letters, parcels, and the like; postal system	87.68	flexible armour made of overlapped or linked metal rings	12.32	0.85	1.53
mat	a piece of material, especially a strong carpet fabric such as wool, that is used as a covering	80.58	a piece of cardboard or other stiff material that acts as a frame for a picture	17.56	0.82	3.80
mate	a marriage partner; husband or wife	82.10	in chess, the placing of a king in a check from which there is no escape; checkmate	17.80	0.76	1.90
meal	an occasion when food is prepared and eaten at a specific time	85.93	coarsely ground grain, such as corn	13.57	0.84	3.93
mint	any of a variety of aromatic plants often cultivated as a source of fragrant oils for flavourings such as peppermint or spearmint	67.42	a factory or plant where money is produced under the authority of the government	24.07	0.62	1.47
mortar	a heavy bowl-like receptacle in which substances are ground or pounded into powder	36.28	a bonding substance made from cement or lime, sand, and water	56.39	0.46	3.03
mount	to climb	61.71	a mountain or high hill	37.59	0.55	5.07

novel	a relatively long work of prose	67.79	new and unusual	32.21	0.50	2.40
pack	a container for goods or belongings, usually flexible and carried on the back	82.91	to fraudulently manipulate, as by choosing, for one's own purposes	9.47	0.88	2.20
pad	a piece of soft material used as cushioning to protect from abrasion or impact, such as a cushion, mat, or thick piece of blanket or quilt	73.66	a soft sound of footsteps	20.65	0.70	3.30
palm	the inner surface of the hand, between the wrist and the base of the fingers	66.26	any of numerous mainly tropical evergreen plants	33.79	0.46	1.93
peel	to pull, tear, or cut the outer covering from	90.41	a spade-like instrument with a long handle used for putting goods to be baked into an oven, or for removing them again	8.64	0.91	2.60
peer	a person of the same rank, status, age group, ability, or the like as another person	55.28	to look closely, searchingly, or with difficulty	44.72	0.37	1.97
pen	any of various instruments used for writing or drawing in ink	74.58	a small, usually fenced enclosure for animals	25.12	0.64	1.53
pet	an animal, usually domesticated or tamed, kept in one's home for companionship rather than practical use	91.20	a sulky mood or fit of temper; peeve; petulance	8.40	0.90	1.80
pine	any of numerous evergreen trees that have clustered, needle-shaped leaves and bear cones	63.47	to be affected with great desire and longing (often followed by for)	36.53	0.46	1.57
pink	a pale reddish colour; the colour that results when white and red paint are mixed	93.85	to cut a scalloped or notched edge on	5.80	0.95	1.40
pit	a wide and deep hole dug or existing in the ground	71.82	the hard seed at the centre of an apricot, cherry, or plum	24.00	0.63	2.17

plane	a flat or level surface	33.99	any of various carpentry tools with an adjustable blade, used to make surfaces of wood smooth	26.29	0.52	4.23
plot	the story line or sequence of events in a novel, play, or the like	57.66	a small piece of land, especially one used for a specific purpose	42.04	0.36	1.63
plump	rounded in appearance; somewhat chubby	85.57	to come down quickly and with all one's weight	14.33	0.79	3.80
pool	any small area of liquid that has collected on a surface; puddle; pond	68.32	a grouping of businesses or individuals for the purpose of combining resources and deriving common benefit from their association	26.04	0.62	2.40
pose	to take or hold a bodily position, as in modeling clothing or having one's likeness painted or photograph taken	71.76	to puzzle or embarrass with a difficult problem or question	28.24	0.58	2.10
prop	to support, stabilize, or sustain with or as if with a beam, stick, stone, or the like	48.76	a piece of furniture or other article used for a theatrical presentation or the like; stage property	50.34	0.28	3.33
pulse	the periodic throbbing of arteries that results from the beating of the heart, especially as felt in the wrist or neck	73.77	an edible seed found in the pods of legumes such as peas or lentils	26.03	0.63	1.70
pump	a mechanical or biological device for compressing a fluid or gas, or moving it from one place to another, especially through pipes or the like	69.94	a simple low women's shoe without buckles, laces, or the like	28.96	0.60	1.43
pupil	any person who studies under a teacher	56.62	the small, dark, circular opening in the centre of the eye	43.38	0.30	1.53

rail	a horizontal bar that extends between two posts and serves as a barrier, fence, or the like	82.55	to denounce someone or something in harsh or abusive terms; condemn	17.43	0.77	1.77
rank	the relative position of one person or group of persons to another in a hierarchy or society	63.56	having an extremely offensive odour or taste	36.44	0.53	1.57
ray	a thin beam of radiation, especially light	76.32	any of various fishes having cartilaginous skeletons, flattened bodies, large pectoral fins, and long narrow tails	22.44	0.69	1.50
rear	the back part of something	69.80	to raise to maturity	29.97	0.55	1.87
reef	a ridge of rock, sand, or coral rising to or near the surface of marine waters	86.79	the part of the sail that is drawn in and tied down to decrease the area exposed to the wind	12.82	0.82	2.37
repair	to restore to a sound state or condition following damage or injury	93.15	to go, as to another place	6.85	0.93	2.37
sack	a large bag, usually made of coarsely woven material or thick paper, used for holding or transporting items in bulk, such as feed, gravel, potatoes, and the like	59.80	to rob of valuables after capturing; plunder; despoil	16.45	0.73	2.83
scrap	a small bit or fragment, especially a leftover or discarded piece	63.66	a fight, especially a physical fight; scuffle	36.34	0.49	1.97
seal	a design, emblem, or embossed or impressed figure used to make a document authentic or official	41.37	any of a number of flesh-eating mammals that live in and around the ocean and have flippers instead of legs	58.58	0.44	1.30
settle	to finally agree upon or resolve	89.12	a long wooden seat with arms and a high back; settee	10.88	0.85	1.87
shed	a simple, usually one-story structure used for storage	67.90	to cast off, take off, or let fall (a covering or growth)	31.30	0.52	1.93

sheer	transparently thin or fine	63.96	to swerve or cause to swerve from a course	29.87	0.70	2.10
soil	the uppermost layer of the earth's surface	67.67	to make unclean; dirty	32.33	0.48	4.27
spill	to cause or allow (a liquid or small particles or objects) to flow or fall from a container	93.35	a thin piece of wood or twist of paper used to light a flame	6.65	0.94	1.80
spray	water or another liquid flying or falling in fine droplets, as from the nozzle of a hose	85.87	a single shoot or branch that has leaves, flowers, or berries	13.93	0.85	2.60
squash	to press, beat, or crush into a pulp or flat mass	71.19	a gourd-like fruit that is borne on a vine-like plant and is eaten as a vegetable	28.81	0.56	2.17
stable	fixed, firm, or steady in position; not shaky or easily moved or overturned	55.94	a building, often containing stalls, where domestic animals, especially horses or cows, are kept and fed	44.06	0.32	2.33
stake	a sharpened or pointed post designed to be driven into the ground, as for a marker or support, or part of a fence	46.34	one's economic or emotional share or interest in something	52.51	0.40	2.00
stall	a small enclosed division of a barn or stable designed for lodging a single animal	41.15	to employ evasion or other delaying manoeuvres	56.11	0.48	2.13
stern	firm and uncompromising	76.80	the rear or back part of anything, especially a nautical vessel	23.20	0.69	1.87
stick	a relatively long and thin piece of wood, especially a stem or branch from a tree or shrub	65.63	to pierce or poke with a pointed object; stab	24.58	0.59	4.80
stir	to agitate or mix (a liquid) by making circular motions with a hand or object	94.48	prison	4.72	0.94	1.43
strand	to beach, or leave behind on the shore	34.89	a length of fibres, threads, filaments, wires, or hairs braided or twisted together	65.01	0.44	2.00

strip	to remove the outer covering or clothing from	61.06	a long, narrow piece or area of mostly uniform width	38.94	0.36	2.20
tap	to strike or cause to strike lightly, and often several times	38.29	a device to control the flow of liquid or gas from a pipe or container, or a plug to prevent such a flow; spigot; faucet	59.53	0.34	2.07
temple	a building or place where a deity or deities are worshiped	65.37	the flat area on either side of the head between the forehead and ear	34.63	0.52	1.87
tend	to have an inclination	48.42	to look after	51.58	0.40	3.23
tense	pulled or stretched tightly; taut	61.60	in grammar, a category of verb inflections that indicate time and duration of an action or state, such as past tense, future tense, and present tense	38.30	0.48	2.10
toast	bread that has been sliced and browned in an oven, toaster, or the like	69.86	a call on other people to drink in honour of someone or something, or a short verbal salute preceding this call	29.46	0.53	1.77
toll	to cause (a large bell) to ring or chime, especially with slow and measured strokes	24.12	a charge or tax, usually for passage across a bridge or along a highway or other road	72.14	0.63	1.90
trace	a visible mark or evidence of a past event or of something having been present	89.57	one of two ropes, chains, or straps used to harness a draft animal to a cart, carriage, or the like	10.43	0.91	2.03
utter	to give forth (a sound or words) vocally	54.30	total or complete in degree or extent; absolute; unqualified	45.70	0.44	2.53
vault	an arched construction of stone, concrete, brick, or the like that forms a roof or ceiling, as in some churches	48.17	to jump, leap, or spring, especially with the use of the hands or a pole as a support	39.44	0.49	2.07
verse	poetry, or a poem, especially in metrical form	78.48	to make knowledgeable or skilled; train; school	21.52	0.72	2.87

wax	a solid yellowish substance that is pliable when melted and is secreted by bees for constructing their honeycombs; beeswax	86.39	to increase gradually in quantity, strength, volume, or the like	12.59	0.81	1.87
yard	a unit of length equal to three feet or 0.9144 meter	46.64	an open area next to or surrounding a house or other building	53.36	0.44	2.63

Appendix 2: Experiments 2-4: Sets of prime-target word pairs. "HF target" and "LF target" refer to the high-frequency and low-frequency meanings of a homonym, respectively. "Rating1" refers to prime-target relatedness ratings (where 1 denoted "highly unrelated" and 7 denoted "highly related") made in the absence of contextual bias. "Rating2" refers to relatedness ratings made when primed with a sentence referring to the low-frequency meaning of a homonym.

Prime word		Related word pairs					Unrelated word pairs			
		HF target	**Rating$_1$**	**LF target**	**Rating$_1$**	**Rating$_2$**	**Target A**	**Rating$_1$**	**Target B**	**Rating$_1$**
Balanced homonym	bay	creek	5.63	alcove	5.58	5.55	tune	1.88	ride	2.85
	bust	breast	5.65	burst	4.73	5.65	basil	1.65	eat	2.28
	calf	knee	5.78	cattle	6.60	6.55	trench	2.03	bitter	1.73
	camp	tent	6.60	gay	5.65	5.40	lag	2.10	quick	2.05
	fan	cheer	6.51	breeze	6.10	5.95	snake	2.00	cancel	2.25
	forge	advance	5.36	hammer	6.10	6.00	bird	1.50	pig	1.82
	jam	knife	5.60	tight	5.33	5.70	oval	1.70	devil	1.53
	lean	bend	5.15	slim	6.40	6.30	crime	2.15	roar	1.74
	novel	poem	5.60	unique	5.53	5.90	wipe	1.53	reward	3.10
	palm	wrist	6.08	exotic	5.36	4.80	sing	2.28	mile	1.75
	pine	oak	6.50	desire	5.28	4.95	cloak	2.10	stroll	3.05
	plot	writer	6.28	acre	5.98	5.45	curl	2.18	plug	2.33
	prop	pillar	4.95	actor	5.53	5.80	parrot	2.93	dinner	2.33
	pupil	lesson	6.23	lens	5.95	6.00	enter	2.58	pan	1.75
	rank	fifth	5.20	odour	5.60	6.20	device	2.10	rift	2.25
	scrap	pieces	5.73	argue	5.23	5.90	castle	1.88	beach	2.73
	seal	swim	5.15	glue	5.87	5.10	rapid	1.95	monk	1.29
	shed	hut	6.55	skin	6.35	6.30	fight	1.78	dance	2.15
	squash	sports	6.50	potato	5.53	4.55	alive	3.25	anchor	2.35

	stall	delay	6.30	sell	5.69	5.85	lip	1.75	veil	2.20
	strip	naked	6.43	ribbon	5.23	5.25	pond	2.43	eagle	1.95
	tap	sink	6.18	knock	6.44	5.70	beans	1.58	poet	1.85
	temple	chapel	6.13	brow	5.81	5.80	swan	2.00	album	1.56
	tend	habit	5.08	nurse	5.98	5.55	begin	3.00	insect	1.95
	tense	stress	6.62	grammar	5.83	6.20	cook	2.35	tea	1.56
	toast	dish	4.93	beer	5.43	5.40	skull	1.75	ache	1.78
	utter	aloud	5.30	absolute	6.10	5.95	fence	1.58	sister	1.70
	yard	grass	5.78	inch	6.40	6.50	invite	2.68	betray	1.60
Unbalanced homonym	angle	maths	6.23	fisher	5.05	5.90	bronze	2.28	laugh	1.51
	cape	jacket	5.45	ocean	4.74	5.15	error	1.65	mental	1.56
	chord	song	6.23	circle	3.00	5.70	zoo	1.68	sore	1.78
	corn	crop	6.28	toe	2.90	5.60	preach	1.46	quit	2.48
	ear	listen	6.80	cereal	3.87	4.60	shelf	1.73	excess	2.20
	egg	goose	6.18	urge	2.85	6.20	boot	1.60	ankle	2.43
	fleet	navy	6.55	swift	4.90	6.15	smart	2.18	ale	1.83
	flock	herd	6.33	fabric	3.70	5.65	screen	2.08	skill	1.80
	fry	butter	5.55	infant	2.10	5.30	clay	2.13	sign	1.78
	hide	buried	5.55	animal	5.87	6.00	cheap	2.10	acid	2.03
	host	guest	6.35	plenty	3.45	5.85	sand	1.80	throat	1.67
	lock	shut	6.38	comb	3.00	5.10	pest	2.53	saint	2.03
	mate	pal	6.55	chess	5.85	5.60	galaxy	1.55	crust	1.65
	mint	ginger	5.48	coin	5.75	6.30	chin	1.88	mess	1.92
	pad	cloth	4.98	foot	5.13	5.65	anger	2.00	frozen	1.63

	pen	ink	6.78	farmer	4.93	5.05	yeast	1.60	add	2.43
	pit	dig	5.98	cherry	4.58	5.65	gaze	2.23	sting	1.64
	pool	bath	6.08	resource	5.10	5.70	tongue	2.13	blade	1.98
	pulse	vein	6.28	seed	5.34	5.70	milk	2.10	gender	2.18
	pump	flow	5.63	shoes	5.78	6.25	hunt	2.13	jaw	1.79
	rail	barrier	5.95	protest	3.77	5.60	willow	1.95	foam	1.68
	ray	shine	6.20	fish	5.10	6.20	ripe	2.30	coal	2.08
	sheer	thin	5.18	veer	2.67	5.05	fridge	1.73	nose	2.43
	spray	mist	5.98	flower	4.33	4.50	rival	2.15	pigeon	1.90
	stern	strict	6.65	boat	5.95	6.10	gift	1.63	bin	1.78
	toll	levy	5.80	bell	5.08	6.30	focus	2.05	mud	1.93
	verse	poetry	6.43	tutor	4.03	5.65	wet	1.65	jungle	1.60
	wax	warm	4.58	moon	3.85	5.45	dog	1.74	heaven	1.65
Non-homonym Set 1	bald	hairy	6.35	wig	6.35	-	vocal	1.80	ton	1.58
	bulk	huge	6.05	vast	5.15	-	wait	2.05	funny	1.73
	crew	squad	6.35	crowd	5.40	-	arrow	2.00	snow	1.98
	curve	chart	5.18	graph	5.78	-	guard	2.18	flood	2.20
	drain	dry	5.50	liquid	5.85	-	banner	1.68	prince	1.58
	fake	truth	6.25	fraud	6.58	-	expand	1.62	fetch	1.85
	fat	broad	5.73	tiny	5.90	-	click	1.78	witch	2.50
	fee	wage	6.13	permit	5.68	-	mummy	1.85	truce	2.25
	foster	assist	4.93	aid	5.50	-	cash	2.70	sick	2.75
	gap	cavity	5.88	hole	6.18	-	whip	1.90	ward	1.73
	grain	wheat	6.67	rice	6.50	-	fairy	1.88	exit	1.73

	grin	teeth	5.98	glad	5.08	-	folder	1.60	queen	2.30
	heap	stack	6.28	gather	5.00	-	dwarf	1.98	quote	1.83
	hit	shield	5.65	slap	6.48	-	reader	2.10	prefer	2.18
	hook	sharp	6.21	trout	6.23	-	busy	1.93	neck	3.25
	hurdle	bounce	4.80	skip	5.20	-	duke	1.70	echo	1.65
	mask	hat	5.25	hood	5.93	-	tide	1.72	canoe	1.75
	raid	rob	6.40	troops	5.95	-	vase	1.68	clown	1.78
	saddle	pony	6.49	camel	5.55	-	angel	1.65	frown	1.80
	scan	copy	6.48	print	5.83	-	beak	1.78	shout	2.33
	elbow	muscle	5.38	bone	6.54	-	envy	1.67	loud	1.56
	shade	shadow	6.43	tree	5.88	-	kiss	2.28	mug	1.74
	silk	linen	6.33	shiny	5.48	-	cheese	2.38	rage	1.55
	slice	divide	6.05	sword	6.10	-	ghost	1.63	active	2.13
	smash	crush	6.43	grind	5.93	-	worm	1.88	virus	2.03
	tall	giant	6.55	height	6.68	-	code	1.53	worry	1.98
	trim	barber	6.23	beard	6.15	-	bag	2.55	spoon	1.90
	wool	yarn	6.28	goat	5.25	-	bread	2.55	foe	1.90
Non-homonym Set 2	abuse	harm	6.58	cruel	6.69	-	menu	1.45	chalk	1.65
	bet	luck	5.80	gamble	6.80	-	parent	1.80	collar	1.75
	burn	grill	6.03	heat	6.38	-	hint	1.72	famous	2.20
	dawn	dusk	6.53	bright	5.25	-	rebel	1.56	toss	1.54
	deaf	blind	6.08	noise	6.10	-	purse	1.45	golf	1.55
	dip	plunge	5.68	rinse	5.08	-	dragon	1.65	humble	2.00
	drift	wander	6.00	yacht	5.15	-	comedy	1.95	gun	2.03

feast	supper	6.18	cake	5.70	-	smooth	1.59	horn	2.73
fog	cloud	6.28	rain	6.03	-	scream	2.10	hug	1.75
fur	fox	5.93	rabbit	6.20	-	chain	2.50	pill	1.88
grasp	grab	6.50	snatch	6.00	-	melt	2.68	trial	2.48
hay	farm	6.13	nest	5.18	-	pearl	2.23	resist	1.55
honey	sauce	5.50	sweet	6.72	-	fun	2.50	rugby	1.50
leap	runner	4.95	jump	6.67	-	owl	2.15	powder	1.65
load	cargo	6.23	lorry	6.18	-	tour	2.35	rub	1.69
loop	rope	5.65	shape	5.03	-	sniff	1.85	tribe	1.95
peak	hill	6.15	climb	5.98	-	batch	1.93	bug	2.18
pilot	sky	6.23	cabin	6.18	-	dirt	2.15	tape	2.63
push	hurt	4.73	ram	5.68	-	rat	1.54	snack	1.88
ritual	pray	6.13	cult	6.30	-	stew	1.98	honest	2.85
rod	copper	5.18	cane	5.75	-	era	1.55	pillow	1.90
smoke	vapour	6.15	oven	5.93	-	dairy	1.95	twin	1.63
sour	apple	5.58	candy	5.80	-	bullet	1.54	weapon	1.60
spy	agent	6.58	enemy	5.95	-	pale	1.78	toad	1.80
teach	guide	6.28	learn	6.58	-	escape	2.15	edge	2.15
tin	bottle	5.18	metal	6.60	-	sad	1.50	track	2.60
torch	cave	5.23	lamp	6.45	-	speed	2.43	scalp	2.53
void	null	6.45	valid	5.33	-	island	2.58	rural	2.85

Appendix 3: Experiment 5: Rodd et al.'s (2012) paragraphs describing new related meanings for trained prime words (in bold). Primes in adjacent paragraphs were swapped across to create new unrelated meanings.

A **mouse** is a prototype of the latest innovation in car design that was unveiled recently at a car show by its designer at a Chinese based firm. The **mouse** prototype boasts a reduced bonnet, sleek bodywork and a slender overall size that minimises the **mouse**'s spatial dimensions. This enhances performance when moving in and out of narrow inner city streets, with the **mouse** claimed to provide an answer to all inner city urban requirements. The **mouse** has the potential to improve traffic flow and cut down on congestion in busy city centres.	In international trading between countries, one country may refer to another as a **farm**. A country that is considered a **farm** can generate and export products at a cheaper rate than could be produced on home soil. Typically **farm** refers to the exploitation of agriculture in countries of low economic development. However, **farm** has also been used between countries with strong economies. For instance, a country may refer to another country that manufactures technological components or software as a **farm** if that country can generate and export these products at a cheaper rate.
A revolutionary new medical device called a **bandage** has recently been developed. When the **bandage** is fastened to the body it is able to extract blood measures without piercing the skin. At regular intervals, measurements are recorded by the **bandage** and then transmitted to a receiver at the hospital. The **bandage** is particularly useful for eating related disorders and allergies as it can monitor the body's reaction to food intake. The **bandage** has also been implemented in the armed forces where it will identify a soldier's health problems early and ensure swifter medical intervention.	In poker, players make bets during the course of a hand. Gamblers not satisfied with this betting alone may make a side-bet known as a **fee**. A **fee** is made privately among two or more players and is independent of the main game. Mostly players will make **fees** when they are not involved in the current hand. Typical **fees** take the form of bets about what suit or numbers will be shown. However, players have been known to make **fees** on anything, such as what time the next waiter will walk through the door.
No recording device is smaller than the **ant**. The **ant** is virtually undetectable and while it can be hidden, it may even go unnoticed in plain sight. Each **ant** contains a tiny camera that is remote activated and that sends a video feed back to the controller. Ingeniously, the **ant** units are mobile and can be moved around by remote control when they are required to get a better view. However, with the technology comes a high price, which currently limits the use of **ants** to that of government intelligence services.	One American Indian tribe paints a series of lines across the face from ear to ear known as a **path**. The adornment of the **path** is part of an annual event of celebrations. The central line of the **path** varies from brown to orange and symbolises the earth's natural tone. A bordering thin white line is then added to the **path** on females, and a thin black line on males. The painting of the **path** is itself symbolic and at the same time met with reciting of an ancient mantra about dreams.
According to folklore, the **grin** is a mythical monster that walks on two legs like a person. Several stories seem to have emerged after the mysterious disappearance of livestock, which are believed to have been eaten by the **grin**. Sketches found in old fairy-tale books show the **grin** to have a mischievous, fixed smile. The demeanour of the **grin** sometimes misleads people into thinking that it is a friendly creature. However, make no mistake for the **grin** is feared to be a vicious little thing, which you would do best to avoid entirely.	The use of the sociology term **hive** has become increasingly popular in recent years. A family home is referred to as a **hive** when it becomes occupied by at least three generations of the same family. Rather than an easy retirement, the grandparents in a **hive** are often roped into doing household duties. The second-generation in a **hive** become dependent on their parents to play babysitter for their own children while they are out. However, not all is bad in a **hive** as most grandparents undoubtedly relish spending time with their grandchildren.

A new technological feature that can be integrated into mobile phones is the **growl**. The **growl** is a feature that makes a loud noise when the user is in danger. People often report finding it scary walking home alone at night. In such a situation, an individual with **growl** can simply dial a short code. Once entered, a proximity-detector in the **growl** is activated. If someone moves towards the **growl** too quickly from a short distance, the phone signals a loud warning alarm alerting others that the user is in trouble.	**Fog** is the collective term for a group of floating particles that can occur on the inside of the eye. **Fog** particles can sometimes be observed, in particular when looking at a bright light. Specks of **fog** have been known to swoop in front of the retina, almost like a shooting star in a person's peripheral vision. Although **fog** doesn't pose any health risk, a very high number of particles can affect your vision. If **fog** occurs, you should have an eye test as early intervention will prevent it from increasing.
A **widow** is an animal that is forced out of their group. In some species, a weak animal may become a **widow** when it becomes a burden on the survival of the others. Alternatively, when there is a short supply of food, an animal may be turned on by its group and forced to become a **widow**. In species that do create **widows**, animals are almost always alone in expulsion. Creating more than one **widow** can be dangerous for survival by weakening the bonds within groups or even by creating rival groups.	A **stain** is a unique and valuable type of precious stone often used in jewellery. When triggered by a rise in temperature or moisture, the **stain** can dramatically change colour. The appearance of a **stain** can change from a dark purple to a vibrant green, from calm beige to a dazzling turquoise in a matter of seconds. **Stains** vary greatly in size from smaller than a 5 pence piece to larger than a human skull. Superstitious groups have suggested that the colour of a **stain** indicates the mood of nearby spirits.
Cardiac pacemakers are very susceptible to electromagnetic interference. A new biomedical implant known as a **cage** has been invented that can be fitted around the pacemaker. The **cage** protects it from such inferences by acting as a barrier against electrical and magnetic signals. Thanks to the **cage**, people with a pacemaker can now walk safely through security detectors at the airport. The **cage** also allows for a broader range of medical examinations to be conducted. The **cage** will lead to a better quality of life for people who have endured heart problems.	During the aurora borealis or northern lights, you can sometimes see a **pearl**. A **pearl** is a bright ring that appears as the waves of light dance across the sky. The best places to see a **pearl** are in the northern most parts of Canada and Alaska. **Pearls** tend to flicker in and out of focus like a star in the sky. This has meant that it is very difficult to take a photograph of a **pearl**, which has led some to believe that it is nothing more than an optical illusion.
A **crew** is a collection or group of Celtic males that play musical instruments in unison. The music performed by a **crew** is described as a rich, harmonic and layered sound. Players in a **crew** stand in a distinctive free-form formation when performing, which is believed to symbolise the fruitfulness of nature. When a player retires from their position in the **crew**, their closest living relative is expected to take over their position in the group. It is considered to be a great honour to be part of a **crew**.	A **pouch** is the area of land that surrounds where an animal sleeps. Many mammals are known to use a **pouch**, in particular smaller species like mice. Those that use a **pouch** will avoid from foraging in it and will leave it largely undisturbed. The main purpose of the **pouch** is believed to avoid attracting predators to where they sleep, a time when they are at their most vulnerable. Interestingly an artefact of this has remained in humans who avoid causing a disturbance and rarely commit crimes in the **pouch** around their home.
The most illustrious names in cooking discuss the burning questions of the food industry, at a conference known in the trade as the **feast**. The **feast** takes place annually at a resort in the British countryside. The **feast** attracts famous chefs from all over the world. The **feast** conference was intended as a centre for	Environmental concerns are becoming an important consideration of those in the transport industry. To ease these concerns a new carbon fibre shell known as a **carton** has been developed. With a lighter frame the **carton** would cut down on fuel emissions considerably. Although

debate, an opportunity to bring cooking experts from around the world together to share ideas. However, increased media coverage of the **feast** has turned the event into more of a publicity stunt than a genuine conference.	questions were raised over the safety of the initial **carton** prototype, it is hoped that the new **carton** model soon to be developed will be as hard as traditional metal frames. If successful, implementation of the **carton** could be widespread, from cars to aircrafts.
Over time all stars and large extrasolar planets 'die' and leave behind a **bone**. The **bone** is a term used by astronomers to describe the residual inner core that remains. The chemical composition of this **bone** is unique to each celestial body. These **bones** are observable by only the very best telescopes found in the world's top institutions. There is much to be gained from their observation and it is hoped that the study of **bones** might reveal more about the big bang and the origins of the universe.	A **cake** is a term that refers to a suspected food package that is brought into prison. **Cake** serves as a code word and is used by prison wardens and police. The use of a word like **cake** means that prisoners are unaware that their suspected package has been spotted. A **cake** usually contains illicit or prescription drugs that are sold for a high price in prison. **Cakes** have also been known to contain much more discreet items such as hit lists or prohibited information on a fellow prisoner.
The **snake** is type of dance move dating back to centuries. The **snake** was mainly performed as part of the entertainment repertoire of street performers. An individual performing the **snake** will elongate their body and then sway from side to side while they keep their head still. As a part of the **snake**, the performers would also weave in and out of each other rhythmically to the sound of the accompanied musician. Facial expressions during the **snake** involved hissing, poking out the tongue and the occasional biting gesture.	**Soup** is the name given to water when it is in its hottest state. In the atmosphere water boils and evaporates as temperature rises, however when also under extreme pressure **soup** is created. For **soup** the liquid and gaseous phase merges into a special type of fluid that is a mixture between the two states. This **soup** is denser than gas but much lighter than liquid. An interesting property of **soup** is that if it touches a material that can withstand the heat, the material will remain dry after contact.
The **spy** is a type of frog that has an amazing talent. The **spy** is able to block out all background noises and focus, undetected, on the calls of female frogs until he hears one that he likes the sound of. The **spy** has an easy time finding a mate despite living by deafeningly loud fast-flowing mountain streams. The **spy** is able to open and close tubes inside the ear, which in other animals remain constantly open. This system used by the **spy** is already being used as a model for 'intelligent' hearing aids.	After a self-sustained nuclear chain reaction, one waste product that may be produced is known as **foam**. Although **foam** is a hazardous by-product, it poses fairly low health risk to humans. As **foam** also cannot become airborne, it is much easier to contain than other forms of radioactive waste. However, the problem posed by **foam** is its long-life span (half-life) and the long-term consequences of contaminated areas. Land that is exposed to **foam** is drained of nutrients, which has devastating consequences for the regions wildlife.
The **join** is an area of land that is a junction between industrial and agricultural areas. With the increased exodus of businesses to cheaper sub-urban or rural areas a **join** is an important consideration for developers. However, it is important to carefully consider the size of the **join**. If the **join** isn't large enough, pollutants from the industries may have negative effects on the agricultural processes, whereas an excessively large **join** on the other hand may restrict development space and push up the price of the land.	A **sip** is a small amount of data that is extracted from a computer file. The individual **sips** of information can easily be recombined when they have all been extracted. While a **sip** can be extracted by anybody, it is predominantly used in relation to hackers. Extracting data in **sips** is employed to reduce the chance of being caught by security software. **Sips** may also be extracted from multiple computers over a longer period of time, which will make it even harder for them to be traced.
A **dawn** is the name for a type of nightmare or unpleasant dream. These **dawn** dreams tend to occur in the early hours of the morning after a long period of deep sleep. The sensation of a **dawn** is reported as being vivid and very	Perhaps the most bizarre footwear ever seen in the animal kingdom known as a **carpet** belongs to a recently discovered snail species. The **carpet** is a covering of scales that is grown to cover its foot. The "**carpet-**

intense, and a characteristic trait of a **dawn** is for the dreamer to awaken from it with a sudden jolt. This sudden adrenaline rush means that those who experience a **dawn** often find it very hard to get back to sleep.	foot" snails were discovered around deep-sea vents at the bottom of the ocean. They form the **carpet** to protect themselves from the toxic chemicals that are pumped into the water at these vents. As a secondary function the hard **carpet** scales appear to have evolved into a protective shield from predators.
A **slot** refers to a safe that is incorporated into a piece of furniture. Each **slot** is individually handcrafted so that intruders are unable to recognize the chief use of the furniture. In front of the **slot,** there is a disguised wooden panel that can be removed to reveal a key lock. Behind this the **slot** fits into a small cavity, from which it slides out. The disguised location makes the **slot** the perfect safe housing storage system for passports, valuable jewellery and marriage or birth certificates.	A **bruise** is a blurred spot that can be found on a photographic picture. When cameras record images onto a film, **bruises** can occur as a result of over-exposure to sunlight. When the film is developed, the **bruise** will appear as a reddish-purple discolouration. However, when exposed to different wave forms of light, **bruises** have also been created in green and brown. While ruining many family pictures, some artists have been able to manipulate the occurrence of **bruises** to create photographic works of art.
The **heap** is a term that describes a unit of measurement used in cooking. The **heap** is a measure of volume roughly equivalent to about five tablespoons. The **heap** is commonly used for powdered cooking additives such as herbs and spices, or dried stock. However, due to the relatively large quantity of the **heap** it is generally only heard in the bigger, professional kitchens as in restaurants or canteens. Wholesalers to such establishments will often sell and may even package ingredients by the **heap**, in pre-prepared wraps.	Camouflage paint that soldiers use in the desert is known as **rust**. **Rust** was developed by a soldier who had studied chemistry at Oxford University before his service. **Rust** is applied to metallic objects such as weapons, machinery or even cooking utensils that may be detected by the enemy. By coating these items with **rust**, they become undetectable by radar equipment beyond short proximity. **Rust** gives soldiers an additional edge over the enemy and has become a vital tool for survival that has saved many lives.

Appendix 4: Experiment 5: Sets of prime-target word pairs. “Rating” refers to prime-target relatedness ratings (1 denoted “highly unrelated” and 7 denoted “highly related”).

Trained word pairs					Untrained word pairs				
Prime	**Related target**	**Rating**	**Unrelated target**	**Rating**	**Prime**	**Related target**	**Rating**	**Unrelated target**	**Rating**
ant	insect	6.4	cruise	1.4	abuse	alcohol	5.9	layer	2.1
bandage	gauze	6.1	coffee	1.5	actor	cinema	6.4	buffalo	1.8
bone	muscle	6.1	flask	2.4	beak	eagle	6.3	prison	1.5
bruise	injury	6.5	pork	2.1	boat	canoe	6.6	kiss	1.8
cage	zoo	6.4	jacket	1.9	butter	bun	5.7	blouse	1.5
cake	icing	6.5	gorilla	1.5	cliff	coast	6.0	desk	1.4
carpet	rug	6.3	monster	1.9	cod	eel	5.6	toy	1.8
carton	package	6.1	heaven	1.5	creek	stream	6.3	skull	2.0
crew	pilot	6.4	falcon	2.5	demon	angel	6.2	ankle	1.7
dawn	horizon	5.9	ship	2.8	elbow	knee	6.2	priest	2.2
farm	ranch	6.2	throat	1.9	fin	dolphin	6.6	sand	3.0
feast	wedding	5.8	leaf	1.9	flower	lily	6.6	arrow	1.6
fee	wage	5.9	beef	2.2	fur	fox	6.4	basil	2.0
foam	bubble	6.2	axe	1.8	goose	pigeon	5.8	fist	1.6
fog	sky	5.7	boxer	1.6	hat	hood	6.0	skeleton	2.2
grin	frown	6.2	fruit	1.7	hay	barn	6.1	beast	2.8
growl	wolf	6.1	cork	1.7	herd	crowd	5.9	monitor	2.3
mouse	cheese	6.1	calcium	1.9	moon	galaxy	6.1	puppy	1.6
heap	mound	6.3	swan	1.6	ocean	lake	6.3	victory	2.6

hive	honey	6.4	copper	2.2	puddle	pond	6.0	thigh	1.7
join	glue	6.1	apple	1.6	reward	medal	6.2	wasp	1.3
path	forest	5.6	bird	2.9	rod	fish	5.8	lunch	2.5
pearl	gem	5.9	pony	1.8	shield	weapon	6.2	thumb	2.4
pouch	purse	6.1	vision	1.7	silk	satin	6.4	dog	1.7
rust	metal	6.4	cave	2.5	sword	knife	6.4	moth	1.8
sip	juice	6.1	golf	1.4	toe	leg	6.2	noise	2.3
slot	coin	5.9	banana	1.6	torch	lamp	6.2	lion	1.6
snake	venom	6.7	coal	1.6	turkey	chicken	6.1	lens	1.5
soup	dish	6.1	prize	1.7	vein	wrist	6.0	cloak	1.7
spy	agent	6.5	flu	1.6	vote	ballot	6.7	liquid	1.8
stain	mud	5.8	tiger	2.0	wig	scalp	6.0	flute	1.6
widow	funeral	6.3	guard	2.3	wool	sweater	6.2	baker	1.9

Appendix 5: Experiments 6 & 7: Sets of prime words. "Sense1" and "Sense2" refer to the dominant and subordinate interpretations of ambiguous prime words, respectively. "Sense relatedness" was rated on a scale from 1 ("highly unrelated") to 7 ("highly related").

Prime word		**Sense$_1$**	**Sense$_2$**	**Sense relatedness**
Balanced homonym	bark	the sound made by dogs	the covering of the trunk of a tree	1.47
	bat	a wooden implement	an animal	1.67
	lean	to move into a sloping position	thin	2.27
	calf	the back of a leg	a young bovine animal	1.97
	camp	temporary accommodation	effeminate	1.83
	cricket	the insect	a sport	1.53
	fan	an admirer	an apparatus	1.53
	nail	a finger nail	a metal spike	2.53
	organ	an instrument	a part of an organism	1.93
	palm	the inner surface of a hand	a tree	1.63
	plot	the main events of a novel	a piece of ground	1.97
	pound	a unit of weight	to hit	1.70
	pupil	a part of an eye	a schoolchild	1.77
	rash	impetuous	a skin condition	2.07
	ring	a shape/circular object	to make sound/phone	2.07
	rock	mineral material	a type of music	1.93
	seal	a substance used to join things	an animal	1.77
	squash	to crush	a vegetable	2.47
	stable	firmly fixed	a building	2.93
	counter	a shop counter	to oppose	1.57

	stall	to evade/delay	a market stall	1.73
	strip	to remove clothes	a long, narrow piece of cloth	3.07
	tap	a device	to touch	2.10
	temple	a building	a part of the head	1.60
	tick	the sound made by clocks	an animal	1.67
	punch	to hit with a fist	an alcoholic beverage	1.90
	toast	sliced browned bread	drinking in honour	1.60
	yard	ground adjoining a building	a unit of length	2.77
Irregular polyseme	ginger	a spice	a red-haired person	3.90
	valid	reasonable	legally binding	5.87
	summit	the highest point of a mountain	a government meeting	3.87
	marker	a writing implement	a test assessor	4.10
	spoil	to diminish value	to harm the character of a child	4.63
	menu	a list of dishes	a list of commands/functions	5.83
	tissue	body cells	tissue paper	3.27
	shade	a shelter from the sun	a colour	5.57
	tape	a narrow strip of material	to record on video tape	4.30
	switch	to change	an electric device	4.27
	coin	money	to invent a phrase/word	2.87
	abuse	misuse/addiction	offensive behaviour	5.97
	tank	a liquid/gas container	a military vehicle	3.20
	screen	an electronic device	to test for the presence of a disease	4.80
	sick	ill	nauseous	6.50
	chart	an illustration	a list of popular songs	4.97

	hood	a covering for the head	a canopy that removes fumes	3.97
	horn	a part of the animal body	an instrument	3.13
	torch	an ignited piece of wood	a flashlight	5.93
	straw	dried stalks of grain	a plastic tube	4.00
	thread	a strand of fibres	a theme	3.70
	foster	to bring up a child	to promote	5.80
	disc	a round object	a cd	5.60
	bulb	a part of some vegetables	a light bulb	3.00
	circle	a shape	a group of people	4.07
	cycle	a sequence of events	to ride a bike	3.13
	hound	a dog used for hunting	to harass/persecute	3.63
	argue	to reason	to quarrel	5.43
Metaphor	drain	to cause liquid to dry/run out	to exhaust	4.73
	sheep	an animal	an easily-led person	4.47
	eye	a part of the body	the centre of a storm	3.50
	lion	an animal	a brave person	4.77
	pig	an animal	a greedy person	4.97
	monkey	an animal	a mischievous person	4.87
	donkey	an animal	a stupid person	4.47
	grill	a device	to question someone	3.13
	tail	a part of the animal body	to follow someone in secret	3.47
	angel	a spiritual being	a kind, well-behaved person	5.80
	milk	a fluid	to exploit/defraud	3.43
	doll	a toy	a polite, considerate person	4.40

	gem	a precious stone	an outstanding, valuable person	5.67
	giant	a mythical being	huge	6.17
	cult	a religious group	a group of devoted fans	5.67
	parrot	a bird	to repeat	4.90
	pearl	a gem	an outstanding, valuable person	5.50
	breeze	a gentle wind	an easy task	3.80
	goat	an animal	a fool	3.27
	chair	a piece of furniture	a leader of a conference	3.27
	stag	an event/party for males	a male deer	4.30
	mirror	a glass surface	to imitate	5.50
	snake	an animal	a traitor	4.27
	wolf	an animal	a rapacious, voracious person	4.97
	fox	an animal	a sly person	4.83
	gift	present	a talent	3.90
	leg	a part of the body	the supports of a table	5.93
	legion	a division of army	a great number of people	5.27
Metonym	violet	a flower	a colour	5.20
	lamb	an animal	the flesh of the animal	6.53
	walnut	a nut	a tree	6.47
	dust	powder	to remove dust	5.77
	juice	the liquid of fruits	to extract juice	5.97
	bottle	a glass container	the contents of a bottle	6.03
	amber	fossilised resin	a colour	5.00
	lemon	a fruit	a tree	6.40

	orange	a fruit	a colour	6.37
	birch	a tree	the wood of the tree	6.63
	cherry	a fruit	a tree	6.33
	peach	a fruit	the mass of the fruit	5.33
	olive	a fruit	a tree	6.40
	almond	a nut	a tree	6.27
	goose	an animal	the flesh of the animal	6.43
	tin	a metal container	the contents of a tin	6.23
	dish	the contents of a plate	a plate	5.47
	gold	a metal	a colour	6.47
	silver	a metal	a colour	5.93
	tea	a shrub	the dried leaves of the shrub	6.13
	trout	a fish	the flesh of the fish	6.43
	skin	the covering of the body	to remove skin	5.37
	broccoli	a vegetable	the mass of the vegetable	6.47
	bone	a part of the skeleton	to remove bones	5.37
	fig	a fruit	the mass of the fruit	6.13
	apple	a fruit	a tree	6.43
	potato	a vegetable	the mass of the vegetable	6.47
	carrot	a vegetable	the mass of the vegetable	6.43
Unambiguous word	lake	-	-	-
	pub	-	-	-
	pond	-	-	-
	spider	-	-	-

owl	-	-	-
wallet	-	-	-
rugby	-	-	-
guest	-	-	-
planet	-	-	-
barn	-	-	-
ale	-	-	-
ankle	-	-	-
attic	-	-	-
oven	-	-	-
medal	-	-	-
bra	-	-	-
butler	-	-	-
bullet	-	-	-
baker	-	-	-
owner	-	-	-
farmer	-	-	-
fun	-	-	-
mayor	-	-	-
sofa	-	-	-
theft	-	-	-
shirt	-	-	-
tennis	-	-	-
winner	-	-	-

Appendix 6: Experiments 6 & 7: Sets of prime-target word pairs. “Rating1” refers to prime-target relatedness ratings (where 1 denoted “highly unrelated” and 7 denoted “highly related”) made in the absence of contextual bias. “Rating2” refers to relatedness ratings made when primed with a sentence referring to the subordinate sense of a polysemous word.

Prime word		**Dominant-sense word pairs**		**Subordinate-sense word pairs**			**Unrelated word pairs**	
		Target word	**Rating$_1$**	**Target word**	**Rating$_1$**	**Rating$_2$**	**Target word**	**Rating$_1$**
Balanced homonym	bark	dog	6.69	leaf	5.76	-	rum	1.83
	bat	mallet	5.77	wing	6.34	-	galaxy	1.48
	lean	incline	5.92	skinny	6.34	-	helmet	1.62
	calf	thigh	6.24	cattle	6.72	-	cruise	1.59
	camp	tent	6.76	gay	5.45	-	boxer	2.34
	cricket	moth	5.14	polo	5.52	-	rob	1.45
	fan	cheer	5.55	rotate	5.41	-	jaw	1.69
	nail	thumb	6.24	metal	6.07	-	hunt	1.93
	organ	guitar	5.38	stomach	5.83	-	bitter	1.90
	palm	sweaty	5.90	exotic	5.45	-	prison	1.45
	plot	writer	6.14	acre	5.76	-	collar	1.38
	pound	weigh	6.34	smash	5.07	-	captain	1.45
	pupil	lens	5.90	lesson	6.55	-	choke	1.79
	rash	hasty	5.38	blister	5.86	-	image	1.83
	ring	propose	6.07	bell	6.66	-	celery	1.97
	rock	coal	5.72	punk	6.55	-	cabbage	1.62
	seal	glue	5.41	mammal	6.52	-	acne	1.45
	squash	hit	5.41	pumpkin	6.00	-	select	1.66
	stable	steady	5.76	horse	6.69	-	stew	1.52

	counter	pharmacy	5.93	oppose	6.24	-	dwarf	1.69
	stall	delay	6.21	shop	5.97	-	layer	2.17
	strip	naked	6.34	ribbon	5.28	-	cub	2.66
	tap	sink	6.41	knock	6.17	-	sit	1.66
	temple	pray	6.03	brow	5.03	-	cork	1.59
	tick	clock	6.52	bug	6.14	-	salmon	1.51
	punch	fist	6.45	whisky	5.16	-	fossil	1.59
	toast	burnt	6.07	wedding	5.48	-	vein	1.66
	yard	grass	5.34	inch	6.10	-	bacon	1.79
Irregular polyseme	ginger	spice	6.41	freckle	4.69	5.23	secret	1.62
	valid	logic	5.17	license	5.52	5.93	throat	1.66
	summit	peak	6.76	diplomat	4.17	5.57	slot	1.93
	marker	pencil	6.24	exam	5.97	6.10	echo	1.50
	spoil	ruin	6.34	brat	5.86	5.83	jug	1.72
	menu	lunch	6.62	website	3.56	5.67	fool	1.62
	tissue	muscle	6.41	nose	5.72	6.03	pleasure	1.38
	shade	tan	5.48	tint	5.83	6.00	expire	1.97
	tape	sticker	5.21	video	6.45	6.37	horde	1.62
	switch	swap	6.07	electric	6.38	5.60	eel	1.62
	coin	penny	6.38	invent	3.38	5.97	brain	1.69
	abuse	drug	6.31	hurt	6.34	6.57	fang	2.14
	tank	fuel	6.45	weapon	6.03	6.17	kidney	1.86
	screen	cinema	6.48	assess	4.23	6.10	hill	1.66
	sick	fever	6.55	vomit	6.83	6.83	traitor	2.34
	chart	graph	6.59	song	6.24	5.70	lung	2.21

	hood	hat	5.79	fume	3.62	5.03	actor	2.14
	horn	buffalo	5.17	harp	4.52	6.40	itch	1.90
	torch	flame	6.59	battery	6.03	6.83	vinyl	1.59
	straw	weave	4.55	thirsty	5.31	6.43	clone	1.93
	thread	needle	6.66	theme	4.00	5.97	crowd	1.66
	foster	adopt	6.55	promote	4.83	6.23	slim	1.52
	disc	saucer	5.48	record	6.14	6.33	mackerel	1.48
	bulb	onion	6.14	lamp	6.17	6.40	cunning	1.48
	circle	triangle	5.97	gang	4.00	5.90	blade	2.43
	cycle	sequence	5.45	bike	6.76	6.50	tongue	1.69
	hound	track	4.72	bully	4.90	5.83	device	1.52
	argue	assert	5.69	shout	6.31	6.55	fat	1.83
Metaphor	drain	liquid	5.93	exhaust	5.38	6.07	smile	1.93
	sheep	wool	6.76	naive	4.00	5.17	kiss	1.31
	eye	vision	6.76	storm	5.72	6.10	cabin	1.62
	lion	jaguar	6.34	brave	6.03	6.30	pirate	1.52
	pig	mud	6.14	greed	5.62	6.57	lagoon	1.72
	monkey	gorilla	6.52	rascal	5.45	6.00	bead	1.79
	donkey	pony	6.10	stupid	4.52	6.17	cave	1.55
	grill	patio	4.59	interview	5.10	6.00	ocean	1.83
	tail	fur	5.76	spy	3.55	6.00	diamond	1.93
	angel	demon	6.24	behave	4.76	5.57	forest	1.69
	milk	dairy	6.62	exploit	3.03	6.30	falcon	1.62
	doll	toy	6.69	generous	2.00	5.03	citrus	1.31
	gem	emerald	6.55	treasure	6.45	6.40	gust	1.45

	giant	monster	6.14	tall	6.55	6.60	sight	2.03
	cult	religion	6.14	fad	5.03	5.70	grin	1.59
	parrot	bird	6.69	repeat	5.69	6.47	saddle	2.14
	pearl	oyster	6.55	wisdom	5.59	6.27	ram	1.72
	breeze	windy	6.55	ease	4.21	6.37	claw	1.69
	goat	camel	5.48	silly	5.07	5.90	universe	1.62
	chair	throne	6.31	leader	4.52	6.23	calcium	1.45
	stag	event	5.21	deer	6.48	6.13	finger	1.59
	mirror	vain	5.45	copy	6.03	6.17	tiger	1.41
	snake	bite	6.41	deceit	4.86	5.87	flask	1.66
	wolf	howl	6.62	voracious	4.02	4.86	storage	1.45
	fox	cat	5.59	sly	6.07	6.27	wrist	1.69
	gift	wrap	6.66	genius	4.93	5.83	spine	1.48
	leg	elbow	5.86	bench	4.03	5.20	reward	1.66
	legion	troops	6.34	mob	4.69	6.03	honey	1.66
Metonym	violet	tulip	5.83	purple	6.59	6.37	roar	1.76
	lamb	herd	5.62	pork	6.14	6.17	brick	1.38
	walnut	cracker	6.00	maple	5.48	5.87	swan	1.34
	dust	dirt	6.52	vacuum	6.55	6.60	sail	1.62
	juice	drink	6.59	extract	5.86	6.00	sand	1.78
	bottle	glass	6.52	litre	6.00	6.37	fence	1.86
	amber	necklace	5.31	red	5.66	5.57	stork	1.45
	lemon	sour	6.45	flower	3.79	4.97	vampire	1.45
	orange	squeeze	6.10	yellow	6.14	6.30	zoo	1.66
	birch	oak	6.48	timber	6.38	6.17	alcohol	1.90

	cherry	plum	6.03	blossom	6.21	6.27	crime	1.38
	peach	sweet	5.83	pie	5.97	6.03	magic	1.62
	olive	snack	5.34	tree	6.20	6.17	armour	1.41
	almond	peanut	6.24	beech	4.17	5.60	shoe	1.41
	goose	beak	6.10	butcher	4.83	5.20	corner	1.45
	tin	contain	5.38	sardine	6.31	6.60	wasp	1.45
	dish	dinner	6.34	plate	6.52	6.30	dean	1.45
	gold	bronze	6.34	pink	5.03	5.57	penguin	1.66
	silver	copper	6.17	dye	3.41	4.93	pill	1.72
	tea	shrub	4.41	sugar	6.28	6.60	jacket	2.03
	trout	swim	6.21	fillet	5.54	5.80	dollar	1.59
	skin	sore	5.62	peel	6.14	6.40	bamboo	2.14
	broccoli	plant	6.10	salad	5.28	5.83	preach	1.76
	bone	skull	6.52	slice	3.79	5.13	craze	2.72
	fig	grape	5.86	cake	4.14	5.57	priest	1.55
	apple	banana	6.41	twig	4.38	4.93	driver	1.72
	potato	garlic	5.14	mash	6.55	6.60	heaven	1.48
	carrot	tomato	5.79	soup	6.14	6.37	confirm	1.41
Unambiguous word	lake	jungle	4.55	dive	5.90	-	aid	1.79
	pub	wine	6.34	vodka	6.55	-	elm	1.66
	pond	puddle	5.69	swamp	6.10	-	wipe	1.86
	spider	insect	6.03	web	6.79	-	weep	1.72
	owl	eagle	6.03	nest	5.86	-	pamper	1.45
	wallet	purse	6.52	cash	6.55	-	socket	2.10
	rugby	sport	6.76	hockey	6.17	-	fern	2.14

guest	invite	6.14	welcome	6.03	-	sneeze	1.93
planet	moon	6.45	crater	5.11	-	feud	1.93
barn	hay	6.38	farm	6.41	-	pedal	1.45
ale	beer	6.41	lager	6.79	-	cover	1.55
ankle	knee	6.07	toe	6.21	-	dispute	1.31
attic	roof	6.28	loft	6.79	-	harm	1.90
oven	roast	6.52	pizza	6.48	-	flute	1.72
medal	trophy	6.07	prize	6.59	-	spew	1.66
bra	breast	6.24	blouse	5.22	-	clarinet	1.31
butler	servant	6.52	waiter	6.48	-	wood	1.59
bullet	arrow	5.93	vest	5.69	-	curry	1.31
baker	recipe	6.28	cookie	6.24	-	examine	2.00
owner	mansion	4.86	palace	4.62	-	fork	1.83
farmer	ranch	6.00	tractor	6.69	-	missile	1.59
fun	laugh	6.52	joy	6.59	-	meat	1.83
mayor	govern	6.28	elect	6.31	-	abdomen	1.31
sofa	seat	6.34	pillow	5.69	-	wage	1.48
theft	burglar	6.52	fraud	6.31	-	thorn	1.55
shirt	blazer	5.83	coat	5.93	-	guard	2.10
tennis	golf	6.07	gym	5.66	-	spoon	1.66
winner	loser	6.62	compete	6.31	-	willow	1.48