Cognitive and Neural Mechanisms Underlying Semantic Ambiguity Resolution

Sylvia Vitello

2014

Cognitive, Perceptual and Brain Sciences

University College London

A thesis submitted for the degree of Doctor of Philosophy

Declaration

I, Sylvia Vitello, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The ultimate goal of language comprehension is to obtain meaning. However, this is difficult because many words are semantically ambiguous, mapping onto multiple meanings. Semantic ambiguity resolution has proven a useful tool to investigate language processing in general. However, the majority of research has focused on the initial encounter of an ambiguous word. Less work has investigated the processes occurring *after* an ambiguous word is encountered, when the initially understood meaning needs to be reinterpreted in light of subsequent inconsistent information.

The first part of this thesis investigated the cognitive processes underlying semantic reinterpretation, examining how successful listeners are at this process as well as assessing the time course of suppressing and integrating the contextually inappropriate and appropriate meanings respectively. A semantic relatedness task was employed in which participants listened to ambiguous and unambiguous sentences and decided whether a following visual probe word was related or unrelated to the sentence. The results revealed that listeners are highly effective at reinterpretation but that suppression of the inappropriate meaning is delayed relative to integration of the appropriate meaning. The rest of the thesis examined the neural responses to these sentences by using functional magnetic resonance imaging (fMRI) and Transcranial magnetic stimulation (TMS). The fMRI study demonstrated ambiguity-elevated responses in left inferior frontal gyrus (LIFG) and posterior temporal cortex. These responses were modulated by the frequency of the ambiguous word's meanings, such that activation was greater for sentences with a higher likelihood of reinterpretation. The final study developed a TMS paradigm to

examine whether LIFG is necessary for this process, demonstrating evidence that this region may be important for sentence processing more generally. Together, this thesis has advanced understanding into the cognitive and neural processes underlying semantic reinterpretation that have various implications for models of ambiguity resolution and language comprehension in general.

Acknowledgements

I wish to thank everyone who has helped me through my PhD: family, friends, colleagues and acquaintances. Although you are too many to name here, know that I could not have done this without your support.

I would, however, like to send a very special thanks to Jenni and Joe for your invaluable and endless supervision, advice and guidance. I do not know how you always had so much time to endure my incessant questions.

Table of Contents

Abstract 3

Acknowledgements........ 5

Chapter 1 General Introduction 14

1.1 Semantic Ambiguity Resolution 15

1.2 Cognitive Mechanisms Underlying Semantic Ambiguity Resolution 16

1.3 Neural Mechanisms Underlying Semantic Ambiguity Resolution 24

1.4 Summary of Experimental Work 27

Chapter 2 Semantic Reinterpretation: Its Effectiveness and Time Course........ 28

2.1 Introduction........ 29

2.2 Experiment 1: 100 ms Probe Delay 42

2.3 Experiment 2: 300 ms Probe Delay 64

2.4 Experiment 3: 1 Second Probe Delay 71

2.5 Experiments 1-3 Combined 76

2.6 General Discussion 82

Chapter 3 Neural Responses to Semantically Ambiguous Sentences: An fMRI Study........ 97

3.1 Introduction 98

3.2 Method 134

3.3 Results 143

3.4 Discussion 152

Chapter 4 TMS Investigation of the LIFG’s Role in Ambiguity Resolution 171

4.1 Introduction 172

4.2 Method 183

4.3 Results 187

4.4 Discussion 191

Chapter 5 Follow-up TMS Experiments: Methodological Investigations 203

5.1 Experiment 6: TMS Control Condition and Order of Ambiguity Conditions 204

5.2 Experiment 7: Larger and More Homogeneous Stimulus Set 214

5.3 Experiment 8: Vertex Stimulation versus No-TMS 224

5.4 Experiment 9: 20Hz Stimulation to Vertex vs. No-TMS 230

5.5 Experiment 10: Re-examining the Effects of LIFG Stimulation on Sentence Comprehension 235

Chapter 6 General Discussion 245

6.1 Overview 246

6.2 Cognitive Mechanisms Underlying Semantic Ambiguity Resolution 246

6.3 Neural Mechanisms Underlying Semantic Ambiguity Resolution 248

6.4 Limitations and Future Directions..253

References 260

Appendix 295

List of Tables

Table 2-1: Sentence conditions and their corresponding probe conditions............ 40

Table 2-2: Descriptive statistics (mean (SD)) for properties of the ambiguous and unambiguous target words.. 44

Table 2-3: Descriptive statistics (mean (SD)) for properties of the ambiguous and unambiguous sentences... 45

Table 2-4: Descriptive statistics (mean (SD)) for properties of the probes............. 52

Table 3-1: Descriptive statistics (mean(SD)) for properties of the ambiguous and unambiguous target words..136

Table 3-2: Descriptive statistics (mean(SD)) for properties of the ambiguous and unambiguous sentences...137

Table 3-3: Unambiguous sentences > SCN: peak activations at p<0.05 FWE corrected. Sub-peaks that are more than 8 mm from the main peak are indented.144

Table 3-4: Ambiguous versus unambiguous sentences: peak activations at p<0.05 FWE corrected. Sub-peaks are indented following main peak. 145

Table 3-5: Individual subjects' "Ambiguous > Unambiguous" local maxima nearest to the frontal and temporal group peaks..151

Table 5-1: Descriptive statistics (mean (SD)) for properties of the ambiguous and unambiguous target words and the sentence final words.....................................217

Table 5-2: Descriptive statistics (mean (SD)) for properties of the ambiguous and unambiguous sentences...218

Table 5-3: Descriptive statistics (mean (SD)) for properties of the probes.............219

List of Figures

Figure 2-1: Distribution of single-word dominance scores for the 90 ambiguous words. Dominance score is the proportion of participants who chose the meaning of the ambiguous word used in the complete version of these sentences. 49

Figure 2-2: Mean accuracy (A) and mean RT (B) of all probe conditions (by-subjects). These means are taken from the combined analysis of the three experiments' datasets to facilitate comparisons across experiments. Although the values are slightly different to those in the separate datasets because extra items were excluded, the patterns of means and significances remain the same. Error bars show standard errors adjusted to remove between subjects variance................... 59

Figure 3-1: Left hemisphere regions that have been shown to have greater BOLD signal for semantically ambiguous than unambiguous sentences. 98

Figure 3-2: Unambiguous sentence versus SCN contrast displayed on the mean group structural image. Red represents activation significant at p<.05 FWE corrected and yellow represents activation significant at p<.001 uncorrected. 144

Figure 3-3: (A) Ambiguous versus Unambiguous sentence contrast displayed on the mean group structural image. Red represents activation significant at p<.05 FWE corrected and yellow represents activation significant at p<.001 uncorrected. (B) Effect sizes for the Ambiguous > SCN contrast and Unambiguous > SCN contrast averaged across the LIFG and pLITC ROIs. (C) Effect sizes for the contrasts between each dominance condition and the unambiguous sentence condition averaged across the LIFG and pLITC ROIs. 146

Figure 3-4: Inter-subject variability around the ambiguous vs. unambiguous contrast group peaks displayed on the group mean structural image. Red circles are the group peaks and the blue are individual subjects' peaks, with darker blue circles representing more than one subject. (a) Variability around the LIFG group peak. (b) Variability around the OTS group peak. 150

Figure 4-1: Example of a TMS trial. A spoken sentence is presented, followed by four pulses of TMS, followed by a visual probe word. 186

Figure 4-2: Mean accuracy (% correct) for the five probe conditions performed with LIFG (white) and Vertex (black) stimulation. Error bars show standard errors adjusted to remove between subjects variance...189

Figure 4-3: Mean reaction times for the five probe conditions performed with LIFG (white) and Vertex (black) stimulation. Error bars show standard errors adjusted to remove between subjects variance. 191

Figure 5-1: Mean accuracy (% correct) for the five probe conditions performed with LIFG stimulation (white) and with no TMS (light grey). Error bars show standard errors adjusted to remove between subject variance. 210

Figure 5-2: Mean reaction times for the five probe conditions performed with LIFG stimulation (white) and no TMS (light grey). Error bars show standard errors adjusted to remove between subject variance. 211

Figure 5-3: Mean accuracy (% correct) for the related probes of ambiguous and unambiguous sentences, performed with LIFG (white) and Vertex (black) stimulation. Error bars show standard errors adjusted to remove between subject variance...222

Figure 5-4: Mean reaction times for the related probes of ambiguous and unambiguous sentences, performed with LIFG (white) and Vertex (black) stimulation. Error bars show standard errors adjusted to remove between subject variance...223

Figure 5-5: Mean accuracy (% correct) for the related probes of ambiguous and unambiguous sentences, performed with Vertex (black) and no TMS (grey) stimulation. Error bars show standard errors adjusted to remove between subject variance...227

Figure 5-6: Mean reaction times for the related probes of ambiguous and unambiguous sentences, performed with Vertex (black) and no TMS (grey) stimulation. Error bars show standard errors adjusted to remove between subject variance. 228

Figure 5-7: Example of a TMS trial. A spoken sentence is presented, followed by four pulses of TMS, followed by a visual probe word. The red arrows represent 20 Hz stimulation and the black arrows represent 10 Hz stimulation for comparison. 231

Figure 5-8: Mean accuracy (% correct) for the related probes of ambiguous and unambiguous sentences, performed with 20Hz Vertex TMS (striped) and no TMS (grey) stimulation. Error bars show standard errors adjusted to remove between subject variance. 233

Figure 5-9: Mean reaction times for the related probes of ambiguous and unambiguous sentences, performed with 20 Hz Vertex TMS (striped) and no TMS (grey) stimulation. Error bars show standard errors adjusted to remove between subject variance. 234

Figure 5-10: Mean accuracy (% correct) for the related probes of ambiguous and unambiguous sentences, performed with 20 Hz LIFG TMS (white) and 20 Hz Vertex TMS (striped) stimulation. Error bars show standard error adjusted to remove between subject variance. 237

Figure 5-11: Mean reaction times for the related probes of ambiguous and unambiguous sentences, performed with 20 Hz LIFG TMS (white) and 20 Hz Vertex TMS (striped) stimulation. Error bars show standard error adjusted to remove between subject variance. 238

Chapter 1 General Introduction

1.1 Semantic Ambiguity Resolution

The purpose of language is to convey meaning, enabling the efficient sharing of information, knowledge and experiences. However, understanding the meaning of an utterance is computationally complex on many levels. One difficulty is that many words do not have a single meaning but refer to several, sometimes completely unrelated, concepts, such as “bark” which refers to the sound that a dog makes as well as to the material from a tree trunk. Thus, listeners must frequently use contextual information to discern which meaning is intended by the speaker, while ensuring that their understanding of the word and ongoing discourse is not affected by its alternative, irrelevant, meanings.

Understanding the processes underlying semantic disambiguation is important to develop comprehensive theories of language comprehension for several reasons. Semantic disambiguation is a fundamental part of everyday communication. Semantically ambiguous words are abundant in natural language, representing the rule rather than the exception. Their ubiquity has been confirmed in calculations of the number words in a standard dictionary that have multiple definitions. For example, Rodd et al. (2002) estimated that at least 80% of English words are ambiguous based on entries in the Wordsmyth dictionary (Parks, Ray, & Bland, 1998). Additionally, current words continue to acquire new meanings which will need to be dealt with by the comprehension system (Rodd, Berriman, et al., 2012). For example, “twitter”, “tablet”, “spam”, “window”, “blackberry” and many more have been adopted by novel technological concepts. Moreover, ambiguous words have been of great interest to cognitive psychologists for over forty years (e.g., K. S. Binder & Morris, 2011; Foss, 1970; Leinenger & Rayner, 2013; Mackay, 1966;

Rayner & Duffy, 1986; Rodd, Lopez Cutrin, Kirsch, Millar, & Davis, 2013; Simpson, 1981; Swinney, 1979; Twilley & Dixon, 2000; Vu, Kellas, & Paul, 1998) because these words provide a window into fundamental cognitive processes, such as selection, inhibition and contextual integration. Such research has led to important and influential insights into the language comprehension system, including contributing key evidence in favour of both modular and interactive accounts of lexical access (e.g., Duffy, Morris, & Rayner, 1988; Swinney, 1979; Twilley & Dixon, 2000; Vu et al., 1998), elucidating factors that affect meaning access and integration (e.g., Rayner, Pacht, & Duffy, 1994; Sereno, O'Donnell, & Rayner, 2006; Simpson, 1994; Twilley & Dixon, 2000; Vu et al., 1998) and demonstrating the fluidity of fundamental lexical properties such as meaning frequency (Rodd et al., 2013).

Understanding into the cognitive and neural mechanisms underlying ambiguity resolution has been considerably advanced by examining how ambiguous words are processed within sentences (Simpson, 1994). Sentences provide a window into natural ambiguity resolution since ambiguous words are usually encountered in these contexts. In addition, sentence comprehension not only requires the perception and identification of an ambiguous word, but, critically, also requires the disambiguation of its meaning. The latter is, arguably, the most fundamental part of ambiguity resolution since it is essential for successful communication.

1.2 Cognitive Mechanisms Underlying Semantic Ambiguity Resolution

A wealth of behavioural research converges on several key cognitive processes that underlie both the processing and disambiguation of an ambiguous word during sentence comprehension. These may be broadly categorised into those associated with initial meaning access, initial meaning selection and semantic reinterpretation (Duffy et al., 1988; Rodd, Johnsrude, & Davis, 2010; Simpson, 1994; Twilley & Dixon, 2000).

1.2.1 Initial Meaning Access

Early, strict, “exhaustive access” models of semantic ambiguity resolution argued that when an ambiguous word is encountered during a sentence (e.g., “the man went to the bank…”) all of its meanings are automatically accessed regardless of contextual information or meaning frequency (also known as meaning dominance) (Onifer & Swinney, 1981; Swinney, 1979). This assumption was primarily supported by cross-modal semantic priming studies in which participants made lexical decisions to, or named, visual probe words that followed the ambiguous word (e.g., Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). The consistent finding was that when the probe was presented immediately after the ambiguous word, responses were faster for probes that were semantically related to either meaning of the ambiguous word compared to unrelated probes. Critically, priming occurred even when context preceding the ambiguous word biased its meaning (e.g., “the accountant went to the bank…) and regardless of whether it corresponded to the highly frequent (i.e., dominant) meaning or its less frequent (subordinate) meaning. The finding that neither contextual bias nor frequency affected initial meaning access was interpreted as evidence that accessing multiple meanings is the one of the primary, automatic stages of ambiguity resolution.

However, subsequent priming studies called into question the automaticity of such an exhaustive access process, with the finding that only one meaning of an ambiguous was primed in certain sentence contexts (e.g., Glucksberg, Kreuz, & Rho, 1986; Simpson & Krueger, 1991; Tabossi, 1988; Tabossi, Colombo, & Job, 1987; Tabossi & Zardon, 1993). For example, in a set of experiments, Tabossi and colleagues found that only the contextually appropriate meaning was primed if it was the dominant meaning and if the context was strongly semantically constraining such that it highlighted distinguishing features of that meaning (e.g., "the violent hurricane did not damage the ships which were in the port, one of the best equipped along the coast") (Tabossi, 1988; Tabossi et al., 1987; Tabossi & Zardon, 1993). Eye-movement research further demonstrated that even in neutral contexts multiple meanings may not be activated equally as suggested in exhaustive access models (K. S. Binder & Morris, 1995; Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989). Such studies showed that readers were just as fast at reading ambiguous words than unambiguous words in neutral sentences when the ambiguous word had one more dominant meaning (i.e., biased words such as "bank"). It was only when the ambiguous words' had relatively equally frequent meanings (i.e., balanced words such as "bark") that readers fixated on them longer than unambiguous, or biased ambiguous, words. These results suggested that, for biased words, their alternative meanings were not activated equally in parallel; otherwise these words should have incurred an ambiguity-related reading cost.

Together these sets of results contributed to the development of various influential theories of semantic ambiguity resolution. For example, the reordered access model (Duffy, Kambe, & Rayner, 2001; Duffy et al., 1988) argues that both frequency and contextual bias specifically influence the rate at which the alternative meanings of an ambiguous word are accessed. Under this view, in neutral contexts, biased words do not incur processing costs because dominant meanings become available more quickly than subordinate meanings and thus can be immediately integrated into sentence comprehension like the meanings of unambiguous words. The current general consensus in the ambiguity literature is that the extent to which multiple meanings are accessed depends on a combination of contextual bias, contextual strength and meaning frequency although the weight given to these factors varies across models (e.g., Duffy et al., 2001; MacDonald, Pearlmutter, & Seidenberg, 1994; Simpson, 1994; Twilley & Dixon, 2000).

Although the precise influence of these factors and their interactions on meaning access is still being investigated (e.g., Colbert-Getz & Cook, 2013; Leinenger & Rayner, 2013; Sheridan & Reingold, 2012; Sheridan, Reingold, & Daneman, 2009) there are two combinations of sentence context and meaning frequency that seem to consistently induce multiple meanings to be accessed or, at least, accessed at the same rate. Evidence of exhaustive access, as measured by priming or increased reading times, is typically found for balanced words when they are encountered in neutral contexts (e.g., “the man thought the <u>bark</u> was going to…”). This has been argued to reflect the fact that neither frequency nor contextual information is available to influence initial meaning access (K. S. Binder & Morris, 1995; Duffy et al., 2001; Duffy et al., 1988; Seidenberg et al., 1982; Swinney, 1979). The second case concerns biased words. Both dominant and subordinate meanings of a word are often found to be accessed when the preceding context supports the subordinate meaning (e.g., “the *ecologist* went to the <u>bank</u>…”) (e.g., Colbert-Getz &

Cook, 2013; Duffy et al., 1988; Kambe, Rayner, & Duffy, 2001; Rayner & Duffy, 1986; Sereno et al., 2006; Sheridan et al., 2009; Tabossi et al., 1987; Tabossi & Zardon, 1993). According to the reordered access model, in these sentences, both meanings are simultaneously accessed because the context boosts activation of the subordinate meaning while the higher frequency of the dominant meaning causes it to become available at the same time (Duffy et al., 2001; Duffy et al., 1988).

1.2.2 Initial Meaning Selection

The early priming studies that demonstrated access of multiple meanings also produced evidence suggesting that listeners do not maintain these meanings for long but make a rapid selection within a few hundred milliseconds of encountering an ambiguous word (e.g., Seidenberg et al., 1982; Swinney, 1979; Tanenhaus et al., 1979). This conclusion was based on the finding that priming became restricted to targets related to only one meaning of the word when the target was presented 3 syllables or 200 ms after the ambiguous word. Interestingly, such rapid initial selection was found even for balanced words in neutral contexts, suggesting that listeners do not maintain alternative possible interpretations until disambiguating context is encountered but integrate one meaning into their current understanding of the discourse (Seidenberg et al., 1982; Swinney, 1979). Seidenberg et al. (1982) proposed that such selection may occur because of limited processing capacity. Subsequent eye-movement research, however, demonstrated that initial meaning selection can also be a cognitively demanding process, particularly for balanced words that are encountered before disambiguating information (K. S. Binder & Morris, 1995; Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989).

Such studies found that readers fixated longer on balanced words than biased or unambiguous words when contextual information was delayed until later in the sentence. As well as being interpreted as evidence that multiple meanings are activated for balanced words, this reading cost was also argued to reflect a processing difficulty associated with selecting a meaning when neither frequency nor context can substantially constrain this process. It is possible that the processing cost may actually reflect difficulty in maintaining multiple meanings rather than in making the selection. However, semantic priming studies provide corroborating evidence for the latter, showing that only one meaning of balanced words still shows priming after a few hundred milliseconds of the ambiguous word being presented, suggesting that, even for these words, one meaning is selected relatively quickly (Seidenberg et al., 1982; Swinney, 1979).

Nevertheless, not all researchers support the notion of initial meaning selection in the absence of biasing information, arguing, instead, that multiple meanings of balanced words in particular are maintained in working memory during sentence comprehension (Mason & Just, 2007; Miyake, Just, & Carpenter, 1994). This view has been defended by the finding that (high-span) readers show no difficulty reading disambiguating information that follows an ambiguous word, which was argued to reflect the ready availability of both meanings (Miyake et al., 1994). Yet, this result is open to another interpretation that is still compatible with the initial selection account. High-span readers may show no reading cost merely because they are efficient at reinterpreting the meaning of sentence rather than because they have maintained both meanings (Twilley & Dixon, 2000). Twilley and Dixon (2000) provide support for this hypothesis through various computer simulations of their independent activation model of ambiguity resolution. Their simulations show that when an ambiguous word is encountered, one meaning is quickly selected and, critically, the alternative meaning can be rapidly reactivated if the subsequent context is inconsistent with the initial selection and weighted strongly enough. Another reason to be doubtful that meanings are maintained is the fact that numerous studies do show that both listeners and readers experience processing costs when encountering disambiguating information late in a sentence such as "*fish in this stream"* in the sentence "usually the <u>bank</u> is not the place to start if you want to catch a *fish in this stream"* (henceforth referred to as late-disambiguation sentences). For example, readers fixate longer on these sections in ambiguous sentences than in unambiguous sentences (Duffy et al., 1988; Kambe et al., 2001; Rayner & Duffy, 1986) and listeners are slower at performing a visual task when it is concurrent to hearing disambiguating information during an ambiguous than unambiguous sentence (Rodd, Johnsrude, et al., 2010). Such costs are argued to reflect the premature selection of an incorrect meaning that needs to undergo a cognitively demanding reinterpretation process when the disambiguating information is encountered. Consistent with the initial selection view, processing difficulty is specifically found for contexts that correspond to subordinate meanings as well as for sentences containing balanced words (Duffy et al., 1988; Kambe et al., 2001; Rayner & Duffy, 1986; Rayner & Frazier, 1989). Those are the two types of ambiguous sentences for which the contextually inappropriate meaning is likely to be initially selected on many trials (Duffy et al., 1988; Simpson, 1994; Twilley & Dixon, 2000). For the former, frequency information will strongly bias the incorrect, dominant meaning to be selected. For the latter, the incorrect meaning will be selected on a substantial, but smaller, amount of trials as there is no systematic frequency bias for a particular meaning, causing it to be selected on approximately half the trials.

Together this body of research strongly corroborates the original conclusion from priming studies that a single meaning is initially selected after an ambiguous word is encountered even in the absence of biasing context. The exact fate of the non-selected meaning, however, is uncertain. For example, non-selected meanings may be completely suppressed (Gernsbacher & St John, 2001; MacDonald et al., 1994) or retain a low level of activation (McRae, Spivey-Knowlton, & Tanenhaus, 1998).

1.2.3 Semantic Reinterpretation

As briefly mentioned, one consequence of not waiting for disambiguating information before making a selection is that the selected meaning may turn out to be incorrect. As a result, it will need to be reinterpreted when information is heard that is inconsistent with that interpretation. However, semantic reinterpretation has received substantially little research focus compared to the initial meaning and selection processes. Yet, it is an equally important area of research as it provides a window into the flexibility of semantic representations generated during comprehension and the operations underlying this process.

Research to date provides a converging body of work demonstrating that reinterpretation is a cognitively demanding process, as evinced by various aforementioned behavioural processing costs (e.g., Duffy et al., 1988; Kambe et al., 2001; Rayner & Duffy, 1986; Rodd, Johnsrude, et al., 2010). Researchers generally agree that this cost reflects time required to suppress the incorrect meaning (e.g., finance meaning of “bank”) and (re)activate the contextually appropriate meaning of the ambiguous word (e.g., “river” meaning) (Duffy et al., 2001; Duffy et al., 1988; Rodd, Johnsrude, et al., 2010; Twilley & Dixon, 2000). However, this is the extent of understanding into this process. Amongst the many unknown details, it is unclear what is the time course of meaning reactivation and suppression during semantic reinterpretation or the extent to which listeners can fully revise and update their understanding of an ambiguous word and the sentence to which it occurs. The first part of this thesis (Chapter 2) focused on this under-examined aspect of ambiguity resolution, presenting three behavioural experiments which revealed insightful findings about both the time course and effectiveness of semantic reinterpretation.

1.3 Neural Mechanisms Underlying Semantic Ambiguity Resolution

The long history of psycholinguistic research into semantic ambiguity resolution provides a well-informed cognitive foundation to examine the neural basis underlying this process. As reviewed above, it is clear that understanding the correct meaning of an ambiguous word entails numerous cognitive processes that ultimately requires the successful integration of contextual and lexical sources of information. Thus, understanding the neural circuitry underlying these processes may provide important insights into how high-level semantic processing is instantiated in the brain.

Neural research into semantic ambiguity resolution has greatly advanced in the recent decade due to the substantial development of functional magnetic resonance imaging (fMRI) as a technique that enables the network-level assessment of brain regions engaged during cognitive tasks (e.g., Mason & Just, 2007; Rodd, Davis, &

Johnsrude, 2005; Rodd, Johnsrude, & Davis, 2012; Rodd, Longe, Randall, & Tyler, 2010; Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007). The increasingly consistent finding from these fMRI studies is that processing semantically ambiguous sentences is supported by a relatively specific set of frontal and temporal brain regions. The left inferior frontal gyrus (LIFG) and posterior middle/inferior temporal cortex have particularly been highlighted. A review of this literature is presented in Chapter 3. However, because of the relative infancy of such research, many fundamental aspects of these neural responses are uncertain. It is unclear what the precise functional roles and mechanisms underlying these regions are. To what extent are these regions related to the specific cognitive processes highlighted in psycholinguistic models of ambiguity resolution such as initial meaning access, initial selection and semantic reinterpretation? What is the relationship between the neural basis of ambiguity resolution and sentence comprehension more generally? For example, do semantic ambiguities place extra demand on brain regions already involved in sentence comprehension or do they recruit additional regions? In addition, to date there has been no investigation of the inter-subject variability of neural responses to semantic ambiguity. This is important as these ambiguity responses have been used as a marker of high-level comprehension in individual patients (Coleman et al., 2009; Coleman et al., 2007). The second part of this thesis (Chapter 3) focuses on a number of these questions. An fMRI study is presented that focuses on the neural responses to initial selection and semantic reinterpretation demands by exploiting the psycholinguistic finding that ambiguous words with either two equally frequent meanings or one dominant meaning load differently onto these processes. This study also examines the relationship between ambiguous and unambiguous sentence responses as well as inspected inter-subject variability of the ambiguity-elevated network.

Despite the important insights that can be revealed by fMRI work, these findings are unavoidably limited in their ability to indicate whether the activated brain regions are *causally* involved in the cognitive task. This is because fMRI, like many other neuroimaging techniques, is a correlational measure of brain activity, measuring neural processes that occur during a task. As a result, it cannot distinguish between the responses that are necessary for the task from those that are merely a by-product of the task. Consequently, based on fMRI research alone, it is unclear the extent to which the frontal and temporal neural responses to semantic ambiguities are serving critical roles in the processing and resolution of their meaning. Some preliminary insights into this question come from patients who have suffered brain damage. Such research has primarily examined the effect of frontal lobe damage on semantic ambiguity processing (e.g., Grindrod, 2012; Hagoort, 1993; Milberg, Blumstein, & Dworetzky, 1987; Swaab, Brown, & Hagoort, 2003; Swinney, Zurif, & Nicol, 1989; Vuong & Martin, 2011). The converging finding is that patients differ in their behavioural performance on ambiguity tasks compared to neurologically healthy adults. Although the exact pattern of differences is not consistent across studies, these constitute support that frontal cortex may play a necessary role in normal ambiguity processing.

However, there are many well-known and uncontrollable difficulties of patient research that limit the precise interpretations that can be drawn from their results. For example, brain damage is often large and varies extensively across patients, there is usually no record of patients' pre-morbid ability and behavioural performance may be confounded by cortical reorganisation (Price & Crinion, 2005). Thus, the final part of this thesis (Chapter 4 and 5) uses Transcranial Magnetic Stimulation (TMS) to investigate the causal role of one region of the ambiguity network, the LIFG. Because TMS is a non-invasive technique, it can be used on neurological healthy adults to examine relatively focal regions of interest, overcoming many of the problems associated with patient data (Devlin & Watkins, 2007; Walsh & Cowey, 2000).

1.4 Summary of Experimental Work

In summary, this thesis investigates three aspects of semantic ambiguity resolution. The cognitive process underlying semantic reinterpretation is examined using a novel experimental paradigm that enables assessment of both the time course of meaning activation/suppression and how effective this process is. The second section examines the neural basis of semantic ambiguity resolution by using fMRI. Specifically, this study assesses the functional roles of ambiguity-responsive brain regions, with a focus on initial selection and semantic reinterpretation demands, as well as examines how consistent the ambiguity network is across subjects. Finally, the third section develops a novel TMS paradigm to assess the causal role of the LIFG in semantic reinterpretation.

Chapter 2 Semantic Reinterpretation: Its Effectiveness and Time Course

2.1 Introduction

The majority of cognitive studies on semantic ambiguity resolution have focused on the event of encountering an ambiguous word and how preceding context affects its processing. Such research has provided important insights into several aspects of the language system, including factors that affect meaning access and integration (e.g., Colbert-Getz & Cook, 2013; Duffy et al., 1988; Rayner & Duffy, 1986; Seidenberg et al., 1982; Sereno et al., 2006; Simpson, 1981; Swinney, 1979; Tabossi, 1988; Twilley & Dixon, 2000; Vu et al., 1998) and shown evidence of the fluid nature of lexical properties (Rodd et al., 2013). Considerably less work, however, has examined the processes occurring *after* an ambiguous word is encountered, when a listener or reader needs to reinterpret their understanding of the word upon encountering contextual information that is inconsistent with their initial interpretation of its meaning. This area of research is important for developing comprehensive models of language as it may reveal the extent to which semantic representations generated during comprehension can be flexibly updated and the operations underlying this process.

2.1.1 Current Account of Semantic Reinterpretation

Semantic reinterpretation is an underspecified feature of models of semantic ambiguity resolution. This is surprising considering that various influential models, such as exhaustive access theories (e.g., Onifer & Swinney, 1981; Seidenberg et al., 1982; Swinney, 1979), reordered access model (Duffy et al., 2001; Duffy et al., 1988) explicitly assume that a single meaning is rapidly selected after encountering an ambiguous word even in the absence of disambiguating information. The selection assumption necessarily implies that an incorrect meaning may be prematurely selected during comprehension that will need reinterpretation if it turns out to be inappropriate. Specifying this reinterpretation process is, furthermore, important because initial selection seems to be a fundamental part of semantic ambiguity resolution during sentence contexts. As discussed in the general introduction, numerous behavioural studies have found that even for words with equally frequent meanings (e.g., “bark”) only one interpretation is favoured in neutral contexts (e.g., K. S. Binder & Morris, 1995; Duffy et al., 1988; Rayner & Duffy, 1986; Seidenberg et al., 1982; Swinney, 1979). Such a selection process may be the result of a limited processing capacity to store alternative interpretations (Seidenberg et al., 1982) and/or may be the result of a highly efficient system that integrates and weights sources of information with the effect that a single interpretation is maintained higher than alternatives in seemingly “neutral” contexts (MacDonald et al., 1994).

It must be noted that the term “selected” does not necessarily mean that the non-selected meaning becomes fully deactivated. Indeed there is little consensus about the precise fate of the non-selected meaning. Many models do assume full deactivation of the non-preferred meaning due to various mechanisms such as passive decay (Duffy et al., 2001; Duffy et al., 1988; Twilley & Dixon, 2000), active suppression (Gernsbacher & Faust, 1991; Gernsbacher & St John, 2001; Twilley & Dixon, 2000), mutual exclusivity of meanings with lateral inhibition from the selected meaning (MacDonald et al., 1994), or computational constraints that allow only one set of semantic features to be represented (Rodd, Gaskell, & Marslen-Wilson, 2004). Other authors, however, argue that, whilst non-selected meanings are suppressed in the left hemisphere, they are maintained in the right hemisphere in case reinterpretation is needed (e.g., Faust & Chiarello, 1998; Faust & Gernsbacher, 1996). Still other models assume some residual activation may remain (McRae et al., 1998). Despite these differences in perspective, there is agreement that non-preferred meanings lose activation when an alternative is selected, which highlights the need of a reinterpretation mechanism that can later re-activate an incorrectly disregarded meaning. However, the operations underlying this process have received relatively little research attention.

Amongst the various models of semantic ambiguity processing, only the reordered access model (Duffy et al., 1988) attempts to describe semantic reinterpretation. According to this model, it is a "time consuming" process (Duffy et al., 1988 p. 440) that entails both accessing/integrating the contextually appropriate meaning and suppressing the incorrect meaning. For example, when reading the late-disambiguation sentence "usually the <u>bank</u> is not the place to start if you want to catch a *fish in this stream",* readers may first interpret "bank" as a financial institution but will have to suppress this interpretation when encountering the last part of the sentence and integrate the river bank meaning instead. This view is based upon eye-movement research showing that readers require extra time to read disambiguating information in late-disambiguation sentences than in unambiguous sentences and, critically, that this ambiguity cost is largest when the disambiguating information biases the subordinate meaning of a biased ambiguous word (Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989). Biased words are the types of ambiguous words for which eye-movement and priming research strongly converge on the finding that the dominant meaning is selected in the absence of preceding biasing context and, thus, will need reinterpretation when subordinate-biased context is later encountered (e.g., Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Simpson & Burgess, 1985; Simpson & Krueger, 1991). Sentences with balanced ambiguous words also show an ambiguity-related reading cost on the disambiguating region but this is smaller than for subordinate-biased sentences (Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989). This is consistent with the notion that balanced words also require reinterpretation but less often than biased words as the correct meaning will be selected on a larger proportion of the trials than for biased words because frequency cannot systematically drive the selection processes toward one particular meaning (Duffy et al., 2001; Duffy et al., 1988; Seidenberg et al., 1982; Swinney, 1979; Twilley & Dixon, 2000).

Processing costs for late-disambiguation sentences have also been shown for spoken sentences. For example, using a dual-task interference paradigm, Rodd et al. (2010) found that participants were slower to make unrelated upper/lower case judgements about a visual letter when it was presented at the same time as they heard the delayed disambiguating information within a spoken sentence, compared with either an unambiguous sentence or a sentence in which the ambiguity had been resolved earlier in the sentence. This finding indicates that delayed disambiguation places a particularly heavy demand on domain-general cognitive resources. Additionally, recent functional magnetic resonance imaging (fMRI) studies have found that late-disambiguation in visual and auditory sentences elicits greater neural activation than unambiguous or dominant-biased sentences (Mason & Just, 2007; Rodd, Johnsrude, et al., 2012; Zempleni et al., 2007). Together, these findings provide converging evidence that when disambiguating information is delayed in either written or spoken sentences, listeners and readers must engage in cognitively and neurally demanding operations to revise incorrect interpretations.

While the cognitive cost of reinterpretation is fairly well-established, little else is known about this process. Two important, yet, uncertain, aspects are 1) the effectiveness of semantic reinterpretation and 2) the time-course of the two fundamental operations: activation and suppression of the contextually appropriate and inappropriate meanings, respectively.

2.1.2 Effectiveness of Semantic Reinterpretation

Current studies do not provide clear evidence about the effectiveness of this process. Can readers/listeners completely recover from a misinterpretation? Recent research on syntactically ambiguous sentences (e.g., “While Anna dressed the baby that was small and cute spit up on the bed”) shows that syntactic reinterpretation may only be partial (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Patson, Darowski, Moon, & Ferreira, 2009). These studies found that inappropriate semantic representations from a reader’s initial parse are sometimes incorporated in their final interpretation of the sentence (e.g., understanding that the baby spit up on the bed *and incorrectly* that Anna had dressed the baby). These findings diverge from computational models of ambiguity resolution which claim that inappropriate analyses are fully deactivated following disambiguation (e.g., MacDonald et al., 1994). The outcome of semantic ambiguity resolution is less clear. The aforementioned processing costs demonstrate difficulty in reinterpreting the meaning of a sentence but do not indicate whether this has been successful.

Although most studies assess participants' explicit comprehension of the sentences alongside the processing speed, it is usually only included to encourage participants to pay attention to the meaning of the sentence and, thus, are usually simple tasks that only crudely measure comprehension. For example, participants may be asked to repeat or paraphrase the sentences or to answer simple yes/no comprehension questions (Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989). Such tasks are not designed to assess whether participants have understood the correct meaning of the ambiguous word in the sentence. However, there are a number of potential reasons why comprehension of late-disambiguation sentences might not be completely successful. Listeners may be unable to integrate contextual information with the correct meaning of an ambiguous word when it occurs substantially after an ambiguous word due to memory constraints and/or may fail to disengage from an initial interpretation of the sentence if it has been too well-integrated during the course of the sentence.

Despite these potential difficulties, two previous studies have found preliminary evidence that reinterpretation can be completely successful for visually presented ambiguous sentences. Miyake et al. (1994) and Zempleni et al. (2007) systematically measured participants' understanding of late-disambiguation sentences using semantic verification tasks. Both studies found that readers were highly successful in integrating the delayed context to understand the contextually appropriate meaning of the ambiguous word and in also using it to reject the contextually inappropriate meaning of these words. However, as both studies were conducted on visually presented sentences, it is unclear whether this is also the case for spoken sentences. Delaying contextual information may have different effects on spoken sentence comprehension. For example, the transient nature of the auditory signal may make integrating such information with the earlier ambiguous word more demanding on memory processes and, thus, reinterpretation may be less successful. In addition, Zempleni et al. (2007) did not measure comprehension when participants first read the sentences. Their participants initially read the sentences in an fMRI scanning session and then performed the comprehension task in a post-scanning session where they re-read the sentences. Thus, it is uncertain how well participants comprehended these sentences on their first encounter.

2.1.3 Time Course of Reinterpretation Processes

The second uncertain aspect concerns the time course of the two key processes of semantic reinterpretation: (i) suppression of the initially selected meaning and (ii) (re)accessing and integrating the previously non-preferred meaning. Existing experimental paradigms that have been used to examine semantic reinterpretation can only show general behavioural costs of disambiguation (e.g., longer reading times or slower performance on concurrent tasks). However, it is also important to investigate these processes separately because different models of semantic ambiguity resolution make different claims about their relative time courses. Various models assume that alternative meanings are mutually exclusive and, thus, suggest that these reinterpretation processes are tightly coupled. For example, in Rodd et al.'s (2004) distributed connectionist model, meanings are represented across a single set of distributed semantic features, which means that increases in the activation of one meaning necessarily correspond to a decrease in the alternative meaning. The opposite perspective is that there is no direct coupling between the activation of alternative meanings, and, thus, the suppression of inappropriate meanings may dissociate from the activation of the appropriate meaning. One clear example is the reordered access model of ambiguity resolution (Duffy et al., 2001; Duffy et al., 1988) which assumes that different meanings are affected by different constraints. According to this model, contextual information only enhances the appropriate meaning without affecting the inappropriate meaning; the inappropriate meaning merely passively decays in the absence of supporting information. This suggests that inappropriate meanings may have a slower time course of being disregarded compared to the time it takes to integrate the contextually appropriate meaning.

Between these two extremes are a number of other influential models which suggest that both patterns (coupling or dissociation of these processes) are possibilities as meaning activation is viewed as being dependent on the combination of numerous sources of information during sentence comprehension. For example, MacDonald et al.'s (1994) probabilistic constraint-based theory assumes that all meanings are directly affected by multiple constraints in parallel (e.g., lexical, semantic, syntactic, pragmatic) and also implements lateral inhibition between meanings. However, they also assume that different meanings may be differently affected by these constraints, depending on their weighting and the current activation levels of the meanings. In terms of semantic reinterpretation, the disambiguating information is critical to understand the correct meaning of the ambiguous word. Thus, the activation of the contextually appropriate and inappropriate meanings will be tightly coupled if the disambiguating information is weighted such that it affects these meanings equivalently to lead to a simultaneous increase in the former and decrease of the latter. However, it is also possible that such information may affect one meaning more quickly than the other since the alternative meanings have different activation levels when the disambiguating information is encountered (i.e., the more frequent, inappropriate, meaning is more active than the less frequent, appropriate, meaning) and/or because the disambiguating information may constrain these meanings to different extents. Another theory that supports alternative possible time courses is the structure-building framework (Gernsbacher, 1990; Gernsbacher & St John, 2001). This model focuses on the influence of contextual information in the enhancement and suppression of appropriate and inappropriate meanings, respectively. Recent versions of this model (Gernsbacher & St John, 2001), in contrast to all aforementioned theories, propose that suppression is specifically transmitted by sentence-level representations rather than by inhibition from the alternative meanings per se. Thus, this theory suggests that, during semantic reinterpretation, the suppression of inappropriate meanings may be delayed until the correct meaning is incorporated into a new interpretation of the sentence. However, as with the constraint-based theories, the extent of this potential delay is dependent on the strength of these representations and the current level of activation of the inappropriate meaning. For example, suppression of the inappropriate meaning may occur instantaneously to the integration of the contextually appropriate meaning if the sentence-level representation is sufficiently constraining. Thus, these models make different predictions about the activation of appropriate and inappropriate meanings during semantic reinterpretation which have not yet been empirically tested. Assessing their time-course may, therefore, reveal important insights into the influence of different constraints on the processes underlying semantic reinterpretation.

2.1.4 The Current Study

The current study is the first comprehensive investigation of the reinterpretation of semantically ambiguous sentence, examining both the effectiveness of semantic reinterpretation as well as the time course of activating and suppressing the appropriate and inappropriate meaning, respectively. For this, a modified version of the semantic relatedness task developed by Gernsbacher, Varner, and Faust (1990) to study the initial processing of ambiguous words was employed. In this task probe words are presented after ambiguous and unambiguous sentences and participants must decide whether they are related to the meaning of the sentence. The processes underlying ambiguity resolution are evaluated by varying the semantic relationship between the probe word and the meaning of the ambiguous word in the sentence. For example, in the original version, Gernsbacher (1990) presented participants with sentences that ended with an ambiguous word (or unambiguous control word). The critical manipulation was that when the probe (e.g., “ace”) was unrelated to the ambiguous sentence (e.g., “he dug with the spade”), it was actually associated with the contextually inappropriate meaning of the ambiguous word (i.e., the card suit meaning of “spade”). Activation of the inappropriate meaning was assessed by comparing how much longer participants took to decide that these inappropriate probes were unrelated to the ambiguous sentence than unrelated to an unambiguous sentence (e.g., “he dug with the shovel”): slower decisions for the ambiguous sentences indicated that inappropriate meanings were more active than baseline. Activation of the contextually appropriate meaning can similarly be assessed by examining reaction times (RTs) for probes related to the contextually appropriate meaning of the ambiguous word (Gernsbacher & Faust, 1991;

Gernsbacher & Robertson, 1995). Furthermore because this task requires an explicit semantic decision, the effectiveness of disambiguation can be indexed by how accurate participants are at accepting contextually appropriate probes as related to the sentence and rejecting inappropriate probes as unrelated to the sentence.

Whilst this paradigm has been used fruitfully to assess the processing of sentence-final ambiguous words, it has not yet been employed to assess the reinterpretation of previously heard ambiguous words. To do this, three changes were made to the original paradigm. First, late-disambiguation sentences were employed that ended with a disambiguating word (e.g., “the teacher explained that the bark was going to be very *damp”*) so that probes could be presented directly after the disambiguating information rather than after the ambiguous word. Second five probe conditions were employed: three for the ambiguous sentences and two for the unambiguous sentences. For the ambiguous sentences, probes were 1) related to the contextually appropriate meaning (“wood”), 2) related to the contextually inappropriate meaning (“sound”) or 3) completely unrelated to the sentence (“pound”). For the unambiguous sentences, probes were either related or unrelated, serving as control conditions. See Table 2-1 for example sentences and probes. The majority of previous studies (e.g., Gernsbacher & Faust, 1991; Gernsbacher et al., 1990) have only assessed performance for either the ambiguous-inappropriate probes or the ambiguous-related probes in a single experiment (see Gernsbacher & Robertson, 1995 for a study that examined both probe types). This current study importantly contrasts all these conditions within the same experiment, allowing us to jointly assess the extent to which contextually appropriate meanings (as indexed by ambiguous-related probes) and inappropriate meanings (as indexed by ambiguous-inappropriate probes) are integrated and suppressed, respectively. The third critical change was our choice of baseline for the ambiguous-inappropriate probe. To assess the activation level of inappropriate meanings, prior studies (e.g., Gernsbacher & Faust, 1991; Gernsbacher et al., 1990) compared the ambiguous-inappropriate probe condition with the unambiguous-unrelated condition. However, the unambiguous-unrelated probe does not take into account general processing costs of comprehending late-disambiguation sentences that is emphasised by previous research (e.g., Duffy et al., 1988; Rodd, Johnsrude, et al., 2010) For this reason, we used the *ambiguous*-unrelated probe as the comparison condition, reasoning that any additional slow-down for the inappropriate probe can be more confidently associated with how active the inappropriate meaning is.

Table 2-1: Sentence conditions and their corresponding probe conditions

Sentence type	**Visual probe word**
Ambiguous (e.g., the woman had to make the toast with a very old *microphone*)	**Related** (e.g., speech)
	Inappropriate (e.g., bread)
	Unrelated (e.g., blanket)
Unambiguous (e.g., the student had to wrap the wrist with a very old bandage)	**Related** (e.g., arm)
	Unrelated (e.g., rock)

Note. In each example sentence, the underlined word is the ambiguous (or matched unambiguous word) and the italicized word is the disambiguating word.

This extended paradigm allows us to test several alternative predictions made by models of ambiguity resolution. First, if contextually appropriate meanings are successfully retrieved and inappropriate meanings are successfully disregarded then participants will have high accuracy for ambiguous-related and inappropriate probes, respectively. Alternatively, reinterpretation may be only partial such that listeners understand the contextually appropriate meaning (hence ambiguous-related probes would have high accuracy) but they may also accept the contextually inappropriate meaning, resulting in low accuracy for ambiguous-inappropriate probes (cf. Christianson et al., 2001). In addition, whilst accuracy can measure the outcome of disambiguation, RTs can provide insights into the processes of disambiguation. First, responses may be slower for *all* ambiguous sentence probes compared with unambiguous sentence probes, suggesting a general processing cost or residual uncertainty about the meaning of the ambiguous sentences. Alternatively we may see a more specific RT cost for the ambiguous-related probes (relative to the unambiguous-related probes), which would suggest the contextually appropriate meaning has not been fully activated/integrated. Similarly, we may see a specific RT cost for the ambiguous-inappropriate probes (relative to the ambiguous-unrelated probes), which would suggest that the contextually inappropriate meaning has not been fully suppressed. Furthermore, the theories of ambiguity resolution that claim mutual dependency between activations of alternative meanings would predict that processing costs associated with integrating appropriate meanings should co-occur with those associated with suppressing inappropriate meanings. Hence, RT costs for related probes should co-occur with RT costs for the inappropriate probes. In contrast, theories that do not explicitly assume such dependency may predict that inappropriate meanings may take more time to be suppressed than it takes appropriate meanings to be integrated and, hence, RT costs for inappropriate probes may be found in the absence of RT costs for related probes.

2.2 Experiment 1: 100 ms Probe Delay

As the time course of semantic reinterpretation is unknown, Experiment 1 tested these predictions using a short probe delay, presenting the probe 100 ms after the disambiguating word.

2.2.1 Method

Participants

25 (15 female) native British English speakers took part in this experiment. Their mean age was 22 years (range: 18-34), all were born and raised in Britain, right-handed and had no known reading or language impairment. Participants were recruited via the University College London (UCL) subject pool, UCL board notices and student mailing lists. Participants were paid for their participation, except first year UCL psychology undergraduates who were given course credit.

Stimuli

(i) Sentences

90 ambiguous and 90 unambiguous sentences were created. On average, the sentences had 12.5 words (SD = 1.24), 16.6 syllables (SD = 1.85) and lasted 2.97 seconds (SD = 0.30). Each ambiguous sentence contained one ambiguous word that occurred, on average, 6.70 words (SD = 1.00) into the sentence. The disambiguating information was provided by the sentence-final word. Four sentences were disambiguated by a compound noun, comprised of two words. This

was necessary to maintain a natural sounding sentence. The disambiguation occurred 4-10 words after the ambiguous word (M = 5.79, SD = 1.46); for example, "the woman had to make the toast with a very old *microphone*", where "toast" is the ambiguous word (i.e., toast refers to grilled bread or a speech given to celebrate an event/person) and "microphone" is the *disambiguating word.* The minimum gap of four words between the ambiguity and the disambiguation ensured that listeners would have selected a meaning (usually the dominant) before hearing the disambiguating information, since research converges on a short window (within about 200 ms) in which selection takes place (e.g., Rodd, Johnsrude, et al., 2010; Swinney, 1979). The ambiguous words were all used as nouns. Some of the ambiguous words possessed two meanings that corresponded to the same spelling (e.g., bark), while others had different spellings (e.g., night/knight). Most of the sentences were modified versions of those used in Rodd et al. (2010), altered such that each ambiguous word was disambiguated by the last word only. Furthermore, to ensure that the words between the ambiguity and the disambiguation did not provide strong cues as to which meaning was more likely to be correct, the sentences were created such that only the last word would need to be changed to disambiguate the ambiguous word to its alternative meaning. For example, in the toast example, "grill" could replace "microphone" for the bread meaning. (Note that the alternative version of each sentence was never included in this experiment.)

A well-matched unambiguous sentence was created for each ambiguous sentence, which had the same syntactic structure but with a low-ambiguity noun (unambiguous target word) in the same position as the high-ambiguity noun (ambiguous target word) in the ambiguous sentence. The ambiguous targets had a significantly higher mean number of meanings (t(89) = 8.20, p < 0.001) and senses (t(89) = 8.23, p < 0.001) than the unambiguous targets, as indexed in the Online Wordsmyth English Dictionary-Thesaurus (Parks et al., 1998). There was no significant difference between the ambiguous and unambiguous targets in terms of word frequency in the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995) (t(89) = 1.55, p = .13), log-transformed word frequency (t(89) = .37, p = .71) or number of letters (t(89) = .52, p = .61). See Table **2-2** for descriptive statistics of these lexical properties.

Table 2-2: Descriptive statistics (mean (SD)) for properties of the ambiguous and unambiguous target words

Target Word	Frequency per million	Log frequency	Number of letters	Number of meanings	Number of senses
Ambiguous	62.3 (79.2)	3.62 (1.02)	4.71 (1.17)	1.94 (0.90)	10.1 (5.66)
Unambiguous	56.69 (63.2)	3.64 (0.94)	4.77 (1.09)	1.09 (0.32)	4.86 (3.10)

The unambiguous and ambiguous sentences did not differ statistically in terms of physical duration (t(89) = .118, p = 0.91), number of syllables (t(89) = 1.62, p = .11) or number of words in the whole sentence. Naturalness ratings were also obtained for the sentences by asking 15 participants who did not take part in the main experiment to rate all sentences on a scale of 1-7, with 7 being highly natural and 1 highly unnatural. Both sets of sentences had relatively high mean naturalness ratings but the ambiguous sentences were rated as significantly less natural than the unambiguous sentences (t(89) = 3.81, p < .001). See Table 2-3 for descriptive statistics of these sentence properties.

Table 2-3: Descriptive statistics (mean (SD)) for properties of the ambiguous and unambiguous sentences

Sentence	Length in seconds	Number of syllables	Number of words	Naturalness rating
Ambiguous	2.97 (0.30)	16.6 (1.85)	12.5 (1.24)	5.47 (0.62)
Unambiguous	2.97 (0.31)	16.4 (1.91)	12.5 (1.24)	5.80 (0.61)

(ii) Meaning Dominance

The ambiguous sentences were constructed to elicit semantic reinterpretation by selecting disambiguating words that corresponded to the less frequent meaning of the ambiguous word as indicated by pre-test scores for a similar set of sentences used by Rodd et al. (2010). To confirm the validity of these meaning preferences, two types of meaning dominance scores were obtained from two independent groups of participants: 1) “single-word” scores, which are participants’ preferences for the meanings of the ambiguous words when they are heard in isolation, and 2) “in-sentence” scores, which are meaning preferences at the point when these words are heard in the experimental sentences just prior to the disambiguation word (e.g., before “supportive” in the sentence “the businessman told him that the <u>fan</u> had been very supportive“). While the “single-word” scores provide an index of the baseline relative frequencies of the alternative meanings of an ambiguous word, the “in-sentence” scores index the extent to which the sentence context influences these preferences. The latter were important to check that, when these words were specifically heard in the sentences used in this study, participants would be likely to prefer the inappropriate meaning prior to the disambiguating word and, thus, need to reinterpret the meaning when hearing that word at the end of the sentence.

54 participants took part in the “single-word” version, all of whom did not take part in the main experiment. Participants performed a modified version of a word association task that is standardly used to measure meaning preferences (e.g. Twilley, Dixon, Taylor, & Clark, 1994). Responses were initially obtained from a set of 15 participants, who performed the task (detailed below) on all the experimental items[1]. The remaining 39 subjects were, subsequently, assessed as part of a larger project that examined a larger set of ambiguous words (N = 406)[2]. Because of this larger number, the stimuli were divided into three sets so that participants only performed a subset of the items. As a result, the ambiguous words relevant to this current experiment were also divided across the three sets, which resulted in each of these participants contributing to approximately 66% of the experimental stimuli. Nevertheless, for all participants, the procedure was as follows. Participants listened to each ambiguous word in isolation (e.g., “fan”) and typed the first related word that came to mind. After participants had performed all word associations, they performed a second task where they explicitly indicated to which meaning of the

---

[1] Three undergraduate students assisted in the data collection.
[2] The data collection for this project was conducted by Dr. Jane Warren.

ambiguous word their word associations corresponded. Participants were informed that each word they had heard had multiple meanings. They then heard each word again and were shown the ambiguous word along with two or three definitions of its meanings. These definitions always included the meaning that was used in the experimental sentence (e.g., “admirer” for the sentence “the businessman told him that the fan had been very supportive”) as well as the “dominant” meaning (e.g., “ventilation device”) and sometimes included a third definition if the word had another highly frequent meaning. They were asked to choose the definition that matched the ambiguous word that they had previously generated. If none fit their interpretation, they were instructed to select the “other” option and write down their own definition. The definition choice allowed participants to code their own word association responses and was included to avoid having to remove unclear word associations that could refer to either meaning of the ambiguous sentence. For example, for the word “fan”, participants may respond with vague words such as “like” or “cool”, which the experimenter would not have been able to code reliably.

All the word associations were coded by the same experimenter (SV) as to whether they were consistent (e.g., “supporter”) or inconsistent (e.g., “wind”) with the interpretation that was used in the whole sentence. When the response was vague, the participant’s definition choice was used to code the response. 7% of responses could not be coded for various reasons including that participants misheard the ambiguous word, did not type in a word association or produced a vague response that was not accompanied by a definition selection. These responses were excluded from subsequent analyses. 5 participants’ total datasets were also excluded from all analyses due to a high rate of missing data or misunderstanding the task.

A dominance score was subsequently calculated as the proportion of included responses (minimum N = 31 per item) that were consistent with the meaning used in the whole sentence. Figure 2-1 shows the distribution of scores. As expected, most sentences had low dominance scores (see Figure 2-1; *M* = 0.25, *SD* = 0.20) indicating that the more frequent, inappropriate, meaning was preferred for the majority of words. Approximately two-thirds of the words had clearly polarised meanings (Dominance $\leq$ 0.3) while approximately one-third had meanings that were relatively equally preferred (0.3 < Dominance < 0.6). The latter words were included in the study as the majority of evidence suggests that, even for such “balanced” words, listeners do not maintain multiple meanings but make a commitment to a single meaning (e.g., Duffy et al., 1988; Rayner & Duffy, 1986; Seidenberg et al., 1982; Simpson, 1994; Swinney, 1979). Thus, reinterpretation will also occur for these sentences if the contextually inappropriate meaning is initially interpreted. These results also showed that for six sentences the more dominant meaning had unintentionally been chosen. These stimuli were also included as they were rare and these meanings were not unanimously preferred, so some participants will still select the incorrect meaning and need to reinterpret the meaning of these sentences.

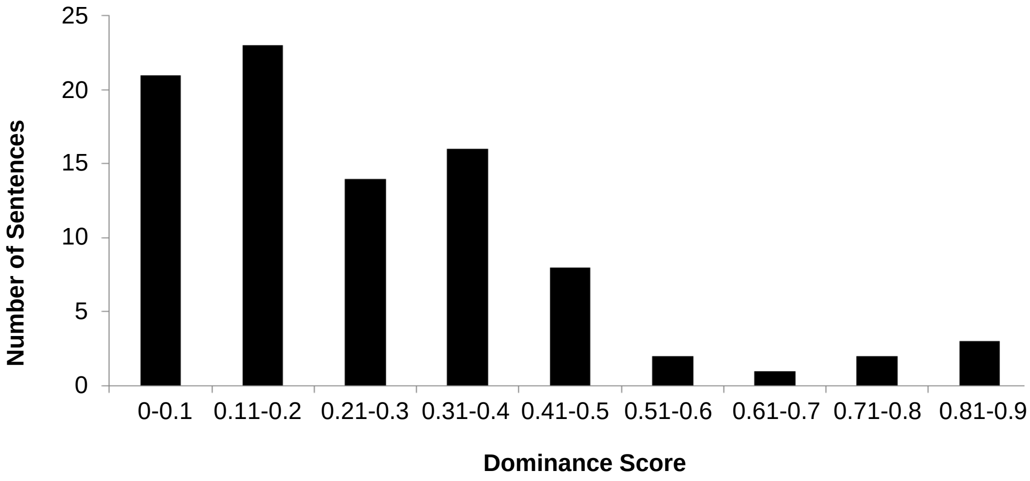

Figure 2-1: Distribution of single-word dominance scores for the 90 ambiguous words. Dominance score is the proportion of participants who chose the meaning of the ambiguous word used in the complete version of these sentences.

An independent group of 30[3] participants performed the "in-sentence" version to examine listeners' preferences for the meanings of the ambiguous words when they are heard in the experimental sentences. It was important that listeners preferred the inappropriate meaning prior to the disambiguating word to ensure that reinterpretation would be induced. The "in-sentence" task was identical to the "single-word" task except that participants generated sentence completions to experimental sentences that had the disambiguating word cut off (e.g., "the businessman told him that the <u>fan</u> had been very"). As for single-word task,

---

[3] Three undergraduate students assisted in the data collection.

participants then indicated the definition of the meaning that corresponded to their completion. The completions were analysed in the same manner as the single-word data, such that responses were coded as being consistent or inconsistent with the meaning of the experimental sentence. Again, a small portion of all responses could not be coded, and therefore were excluded from further analyses (4% of total responses). Two participants' whole datasets were also excluded due to a large number of missing data and/or uninterpretable responses (> 50%). As a result of these exclusions, a minimum of 16 usable data points contributed to each item's "in-sentence" mean dominance scores. These scores was significantly higher than the mean "single-word" score (M = 0.38, SD = 0.25; t(89) = 4.72, p < .001), indicating that, to some extent, the context biased the sentences toward the sentence meaning. However, this absolute score was still low in that the majority of these ambiguous words were still interpreted toward the incorrect meaning in the sentence frames and, thus, would have needed reinterpretation if the disambiguating information had been encountered.

(iii) Probe Words

For each ambiguous sentence, three types of probes were selected: 1) a word related to the contextually appropriate meaning of the sentence (ambiguous-related); 2) a word related to the contextually inappropriate meaning of the ambiguous word (ambiguous-inappropriate), and 3) a word completely unrelated to the sentence (ambiguous-unrelated). For each unambiguous sentence, two types of probes were selected: a word related (unambiguous-related) and a word unrelated (unambiguous-unrelated) to the sentence (see Table 2-1 for examples). Across conditions, there were no significant differences in the probes' length (number of letters: F(4,356) = .44, p = .76), CELEX frequency (F(4,356) = .15, p = .94), log frequency (F(4,356) = .37, p = .83) or Wordsmyth-based number of meanings (F(4,356) = .45, p = .73) and senses (F(4,356) = 1.28, p = .28).

Relatedness ratings were obtained for the probes to assess how semantically related they were to the corresponding sentence. The relatedness task required participants to rate the degree to which the word was related to the preceding sentence on a scale of 1 (very unrelated) to 7 (very related). 34 native British English speakers, who did not take part in the main experiment, performed the relatedness task. Mean relatedness ratings for each probe were calculated by averaging across participants' ratings. Importantly, each related probe had a mean relatedness ≥ 4.5 and a median relatedness ≥ 5. The overall mean rating for the ambiguous and unambiguous related probe conditions were all relatively high (see Table 2-4). However, the ambiguous-related probes had significantly lower ratings than the unambiguous-related probes (t(89) = 2.59, p = .01). Each inappropriate and unrelated probe had a mean relatedness ≤ 2.5 and a median relatedness ≤ 2. There was no significant difference between the ambiguous-unrelated and unambiguous-unrelated probes (t(89) = .64, p = .52). Although the inappropriate probes had a similar low mean rating, this was significantly higher than the two unrelated probes (t(89) = 10.03, p < .001; t(89) = 11.90, p < .001, respectively). Importantly, the finding that the ambiguous-related probes had high relatedness ratings and the ambiguous-inappropriate probes had low ratings, additionally, confirmed that the sentence-final words are effective in disambiguating the meaning of the ambiguous word in the intended way.

Table 2-4: Descriptive statistics (mean (SD)) for properties of the probes

	Ambiguous Sentence Probe			Unambiguous Sentence Probe	
Probe Property	**Related**	**Inappropriate**	**Unrelated**	**Related**	**Unrelated**
Relatedness rating	5.59 (0.64)	1.80 (0.38)	1.27 (0.29)	5.81 (0.55)	1.29 (0.24)
Frequency per million	71.0 (106)	68.5 (95.1)	66.5 (97.9)	62.2 (97.0)	70.7 (100)
Log frequency	3.38 (1.40)	3.48 (1.33)	3.37 (1.37)	3.26 (1.41)	3.38 (1.46)
Number of letters	5.97 (1.79)	5.98 (1.71)	6.03 (1.71)	6.26 (1.73)	6.08 (1.65)
Number of meanings	1.10 (0.37)	1.11 (0.44)	1.09 (0.36)	1.07 (0.25)	1.06 (0.27)
Number of senses	5.73 (4.93)	6.14 (5.77)	5.17 (4.83)	4.87 (3.76)	4.98 (4.56)

In order to create six versions of the experiment (see Design section) the set of ambiguous sentences were divided into three lists, matched (all $p > .1$) on relatedness rating, frequency, log frequency, length, number of meanings and number of senses of each probe condition, and the unambiguous sentences were divided into two matched lists (all $p > .1$).

24 practice sentences were additionally created for use in an initial practice block and 50 filler sentences were created. 20 of these filler sentences were to be

presented at the beginning of the experimental blocks (lead-in sentence) and the other 30 were to be sentences presented during the experiment with a related probe (related filler sentence) to result in an equal number of related and unrelated responses (see Procedure section). The sentences were constructed in the same way as the experimental sentences. Half of each type of filler were ambiguous sentences and the other half unambiguous. All sentences were spoken by the same female speaker[4]. For each practice and lead-in sentence, one probe was selected and probe conditions were of roughly the same proportion as in the experimental task.

Design

Probe condition was manipulated within-subjects so that participants responded to all five conditions. However, to avoid repeating sentences, participants were assigned to one of six versions, such that each participant only encountered each sentence once (i.e., with one probe) but that across participants all combinations of sentence and probes types were presented. Six versions ensured that the assignment of probe type to each sentence was fully counterbalanced because the ambiguous sentence had one more probe condition than the unambiguous sentences. Specifically, the ambiguous sentences had three probe types, requiring three stimulus lists (see Stimuli section) whereas the unambiguous sentences had two probe types, requiring only two stimulus lists, which led to six combinations of the two sets of lists. Thus, in all versions, one third of the ambiguous sentences

---

[4] Dr. Jennifer M. Rodd

were followed by a related probe, one third by an inappropriate probe and the remaining third by an unrelated probe, and half of the unambiguous sentences were followed by a related probe and the other half by an unrelated probe.

Procedure

The experiment was presented on a computer using MATLAB 2010 software (Mathworks, Sherborne, MA, USA). Each trial commenced with a fixation cross presented in the centre of a computer screen. After one second, a sentence was presented through headphones. 100 ms after sentence offset, a probe was presented (font: Helvetica; size: 36pt), replacing the fixation cross. The probe remained on screen until participants responded. Participants were instructed to decide whether the word was related or unrelated to the meaning of the sentence just heard by pressing keyboard button number one for related and number two for unrelated with their index and middle fingers, respectively. Once participants had responded, a blank screen was presented for one second followed by the next trial.

Participants were given written instructions with two example sentences, each with two related and unrelated probes. To ensure that participants treated the ambiguous-inappropriate probes as unrelated to the sentences, one of the examples was of an ambiguous sentence with an ambiguous-inappropriate probe as one of its unrelated probes and participants were informed why this probe was unrelated (i.e., that it does not fit the meaning of any of the words in that sentence). Ambiguity was never explicitly mentioned. All participants completed the same 24 practice trials prior to performing the actual experiment, being given accuracy feedback during the trial and clarification of incorrect answers from the experimenter after the practice.

The trials were divided into five blocks, ensuring that each ambiguous and its matched unambiguous sentence appeared in separate blocks. The blocks were randomised across participants, each starting with four different lead-in sentences, and having 36 experimental sentences and 6 related fillers. All sentences (except the lead-in items) were randomised within each block for each participant. Feedback was not given during the experimental blocks.

2.2.2 Results

Data Preparation and Analysis

One participant’s whole dataset was excluded due to 50% accuracy in the ambiguous-inappropriate probe condition. Sets of items were also excluded if at least one of their probe conditions achieved less than 50% accuracy across participants. A set is comprised of one ambiguous sentence and its matched unambiguous sentence with the five corresponding probe conditions. This led to one set of items being removed due to 38% accuracy for its ambiguous-related probe condition. For this set, the ambiguous word was “cast” and the related probe “plaster”. Individual responses faster than 300 ms or made after the timeout (2500 ms) were also excluded; they contributed to 0.4% of the total data.

By-subject (F_1) and by-item (F_2) means of the accuracy and reaction time (RT) were subjected to two main analyses. In all analyses Sentence Ambiguity and Probe Relatedness were employed as repeated-measures factors. First, 2 × 2 ANOVAs were conducted to evaluate the effects of Sentence Ambiguity (ambiguous vs. unambiguous) and Probe Relatedness (related vs. unrelated) on accuracy rates and

RTs. The ambiguous-inappropriate probe condition was not included in these initial analyses because there was no inappropriate probe condition for the unambiguous sentences. Second, to assess interference of the inappropriate meaning, ANOVAs were conducted that specifically examined the effect of the *unrelated* ambiguous probe type (ambiguous-inappropriate vs. ambiguous-unrelated) on accuracy and RTs. Version was included in all ANOVAs as a dummy variable but main effects and interactions with version are not reported (Pollatsek & Well, 1995). Furthermore, wherever additional analyses were conducted between pairs of conditions, ANOVAs were employed so version could be included as a dummy variable[5].

Accuracy

Overall accuracy was high, with at least 90% correct responses in each condition. However, the ambiguous-related and ambiguous-inappropriate probes had markedly lower accuracy than the other three probe conditions (see Figure 2-2A far left column).

Sentence Ambiguity and Probe Relatedness. This ANOVA showed a significant main effect of probe relatedness, where related probes were less accurate than unrelated probes ($F_1(1,18) = 20.3$, $p < .001$, $\eta^2_p = .530$; $F_2(1,83) = 51.2$, $p < .001$, $\eta^2_p = .381$). There was also a significant main effect of sentence ambiguity, where

---

[5] Preliminary analyses also examined the effects of dominance but these showed no significant effects. These results are not reported further because they may be merely due to low power resulting from a combination of three factors that increase variance: 1) between-item variance due to variation in the sentence-probe relationships, 2) a relatively small number of items in three key dominance condition ($N = \leq 30$; see Chapter 3 for more details) and 3) the mean response values being contributed by only eight participants due to the version manipulation.

ambiguous sentence probes were less accurate than unambiguous probes (F_1(1,18) = 9.60, p = .006, η^2_p= .348; F_2(1,83) = 12.2, p = .001, η^2_p= .128). This main effect of ambiguity was modulated by a significant interaction between sentence ambiguity and probe relatedness such that the ambiguity cost was larger for related probes (5%) than unrelated probes (0.6%) (F_1(1,18) = 8.79, p = .008, η^2_p=.328; F_2(1,83) = 6.01, p = .016, η^2_p= .067). Additional ANOVAs conducted on related and unrelated probes separately revealed that the ambiguity cost was only significant for the related probes (F_1(1,18) = 10.2, p = .005, η^2_p= .361; F_2(1,83) = 9.66, p = .003, η^2_p= .104); there was no significant ambiguity effect for the unrelated probes (F_1(1,18) = 1.13, p = .302, η^2_p= .059; F_2(1,83) = .903, p = .345, η^2_p= .011).

Because the ambiguity effect was specific to the related probes, it was important to rule out the possibility that this was due to the ambiguous-related probes having lower relatedness ratings than the unambiguous-related probes. Thus, a subset analysis was conducted on a set of related probes that were matched on relatedness rating across ambiguity. The subset was obtained by removing ambiguous-related probes with relatedness ratings lower than the minimum ratings for the unambiguous items (4.63) and unambiguous-related probes with high relatedness ratings (> 6.5). 76 out of 90 items were included; the mean relatedness ratings for the ambiguous-related and unambiguous related probes were 5.6 (SD = .63) and 5.7 (SD = .52) respectively, which did not differ statistically (t(75) = .841, p = .403). This confirmed the initial findings: ambiguous-related probes were still found to be less accurate than unambiguous-related probes (F_1(1,18) = 8.03, p = .011, η^2_p= .309; F_2(1,69) = 5.08, p = .027, η^2_p= .069).

Ambiguous Inappropriate Probes. The ANOVA comparing accuracy of the ambiguous inappropriate probes with that of the ambiguous unrelated probes showed a main effect of unrelated probe type, where accuracy was lower for inappropriate probes than unrelated probes ($F_1(1,18) = 34.1$, $p < .001$, $\eta^2_p = .654$; $F_2(1,83) = 32.3$, $p < .001$, $\eta^2_p = .280$).

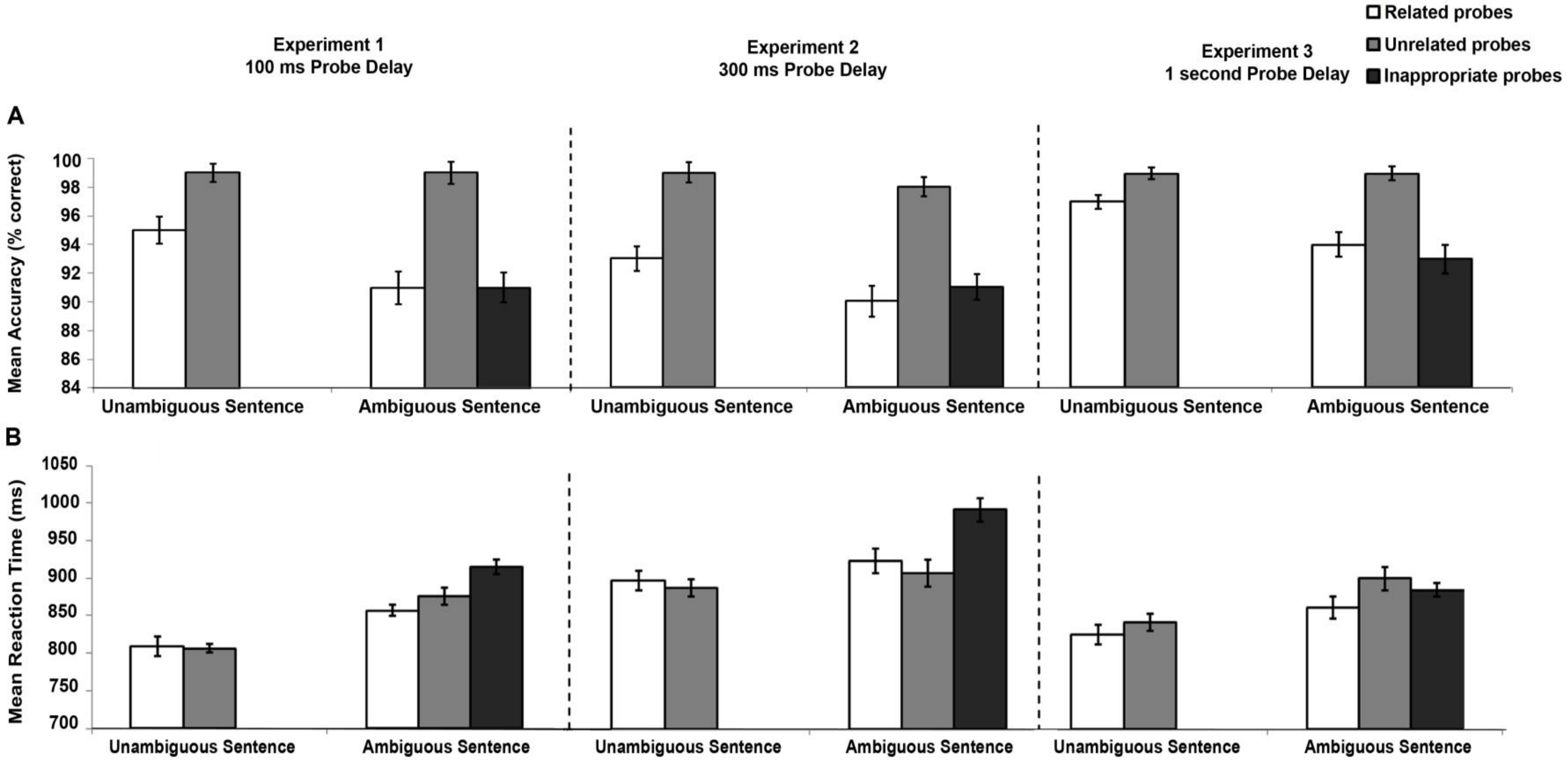

Figure 2-2: Mean accuracy (A) and mean RT (B) of all probe conditions (by-subjects). These means are taken from the combined analysis of the three experiments' datasets to facilitate comparisons across experiments. Although the values are slightly different to those in the separate datasets because extra items were excluded, the patterns of means and significances remain the same. Error bars show standard errors adjusted to remove between subjects variance.

Reaction Times

Only correct RTs were analysed. The same analyses were conducted as for the accuracy data. Inspection of the data revealed a large range of RTs within subjects that resulted from long outliers. To reduce the potential effect of such outliers, the raw RT data were transformed using an inverse transformation. To assess the effect of this transformation on the results, the analyses were re-run on the untransformed data when RTs greater than three standard deviations from each subject's mean had been removed. The results of both analyses were very similar so we report the results of the transformed data and indicate where the patterns of significance differ for the untransformed data. In addition, for ease of interpretation, the means of the untransformed data are presented (see Figure 2-2B far left column).

Sentence Ambiguity and Probe Relatedness. This ANOVA showed a significant main effect of sentence ambiguity, where responses were, on average, 62 ms slower for ambiguous than unambiguous probes ($F_1(1,18) = 25.6$, $p < .001$, $\eta^2_p = .588$; $F_2(1,83) = 23.8$, $p < .001$, $\eta^2_p = .223$). There was no significant main effect of relatedness ($F_1(1,18) = .920$, $p = .350$, $\eta^2_p = .049$; $F_2(1,83) = 1.59$, $p = .211$, $\eta^2_p = .019$) and no significant interaction between sentence ambiguity and probe relatedness ($F_1(1,18) = .485$, $p = .495$, $\eta^2_p = .026$; $F_2(1,83) = .336$, $p = .563$, $\eta^2_p = .004$).

Ambiguous Inappropriate Probes. Comparing the ambiguous inappropriate and ambiguous unrelated probe conditions, the results showed a significant main effect of unrelated probe type, where responses were, on average, 47 ms slower for the inappropriate than unrelated probes ($F_1(1,18) = 12.7$, $p = .002$, $\eta^2_p = .413$; $F_2(1,83) = 8.64$, $p = .004$, $\eta^2_p = .094$).

2.2.3 Discussion

The results revealed interesting, but different, effects within the accuracy and RT data. These two sets of results will be discussed separately because they have different theoretical implications. Whilst accuracy indicates the outcome of disambiguation (i.e., participants' explicit understanding of the sentence), RT, which is analysed for correct trials only, provides insights into the processes that underlie disambiguation when it is successful.

Accuracy

The first important finding was that accuracy was high (M > 90%) for all conditions, including the ambiguous-related and inappropriate probes. As correct decisions for these two conditions require that the contextually appropriate meaning is integrated and the inappropriate meaning is disregarded, respectively, this pattern demonstrates that participants fully disambiguated most sentences, settling on the contextually appropriate interpretation.

Despite high absolute accuracy, performance was not uniform across conditions. In particular, the ambiguous-related and inappropriate probes had relatively lower accuracy. Specifically, although more errors were made for related than unrelated probes (regardless of ambiguity), the results showed a specific difficulty for ambiguous-related probes, indexed by a significant sentence ambiguity by probe relatedness interaction. The ambiguous-unrelated probes, in contrast, had near perfect accuracy that was comparable with the unambiguous-unrelated probes. Importantly, this specific ambiguity cost for related probes cannot be attributed to the difference in relatedness ratings between the ambiguous-related and unambiguous-related probes since it remained in a subset of items that were matched on relatedness. In addition, participants also produced more errors for the ambiguous-inappropriate probes compared with the ambiguous-unrelated probes. This suggests that, on a small, but significant, proportion of trials, participants were unable to use the disambiguating word to either select the correct interpretation of the ambiguous word or to suppress the contextually inappropriate meaning.

One possible reason for these (relatively rare) disambiguation failures is that reinterpretation may have been interrupted by the probe task. Under this view participants would reach the correct interpretation if given more time, which predicts that a longer probe delay should improve accuracy for these two types of ambiguous probes. This prediction is tested in Experiment 2.

Reaction Times

The reaction results showed a rather different pattern. While responses were slower for ambiguous than unambiguous sentence probes, there was no interaction with probe relatedness: related and unrelated probes were similarly delayed by ambiguity. For example, for the sentence “the woman had to make the toast with a very old *microphone”*, participants were slower both to accept “speech” and to reject “blanket”, compared with unambiguous controls. This non-specific slow-down for ambiguous sentences suggests a general processing cost associated with understanding these sentences, presumably reflecting time needed to resolve uncertainty about the correct meaning. This finding is consistent with previous research and shows novel behavioural costs of resolving late-disambiguation sentences (e.g., Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Rodd, Longe, et al., 2010). In addition to this main effect of ambiguity, there was also a specific RT cost for ambiguous-inappropriate probes (e.g., "bread" for the sentence "the woman had to make the toast with a very old *microphone"*) compared to ambiguous-unrelated probes (e.g., "blanket"). As this inappropriate probe cost was found over and above the general ambiguity effect, it suggests that the inappropriate meaning was still more active than unrelated meanings.

Together, these effects suggest that at this short probe delay there is still uncertainty about contextually appropriate meanings and also that inappropriate meanings are over-active compared with unrelated meanings. These results are consistent with descriptions of reinterpretation as a "time-consuming process" (Duffy et al., 1988 p. 440). Importantly both these RT effects are emerging on trials in which participants make a correct response, which suggests that, although participants incur processing costs when resolving late-disambiguation sentences, this does not prevent successful comprehension.

Furthermore, the finding of impairments for both appropriate and inappropriate meanings may provide preliminary support for models claiming that activation of alternative meanings are tightly coupled (e.g., MacDonald et al., 1994; Rodd et al., 2004). For example, the disambiguating information at the end of the sentence did not seem to preferentially help the appropriate meaning, as has been suggested in the reordered access model (Duffy et al., 2001; Duffy et al., 1988), otherwise there should have been an RT cost for the ambiguous-inappropriate probes in the absence of a cost for the ambiguous-related probes. However, it may be possible that such a dissociation between appropriate and inappropriate meanings occurs at a later time when more time has elapsed for constraints to have a differential effect on these meanings. Therefore Experiment 2 examined this possibility by increasing the delay between the end of the sentence and the onset of the probes from 100ms to 300ms. This longer probe delay was chosen because previous research on initial meaning selection suggests that inappropriate meanings can be suppressed by about a few hundred milliseconds after encountering an ambiguous word (e.g., Seidenberg et al., 1982; Swinney, 1979). In addition, this longer delay will allow us to investigate whether the (relatively rare) disambiguation failures seen in the accuracy data are reduced when participants are given more time to process the disambiguating word before the arrival of the probe.

2.3 Experiment 2: 300 ms Probe Delay

2.3.1 Method

Participants

26 (17 female) native British English speakers from the same population as Experiment 1 took part in this experiment. Their mean age was 22 years (range: 18-34).[6]

---

[6] Two undergraduate students assisted in the data collection.

Stimuli, Design and Procedure

All was the same as Experiment 1 except that the probe was presented at 300ms post sentence offset.

2.3.2 Results

Data preparation and analysis

Data were trimmed in the same way as Experiment 1 leading to three participants' data and three sets of items being removed from all subsequent statistical analyses. The excluded item sets had low accuracy in the ambiguous-related probe condition; these probes (ambiguous word) were “layer” (coat), “metal” (cymbal), and “fireplace” (poker). Individual responses faster than 300 ms or made after the timeout (2500 ms) were also excluded; they contributed to 0.8% of the total data. The same analyses were conducted as in Experiment 1.

Accuracy

Like Experiment 1, overall accuracy was above 90%, for all probe types with the lowest accuracy for ambiguous-related and ambiguous-inappropriate probes (see Figure 2-2A middle column).

Sentence Ambiguity and Probe Relatedness. This ANOVA showed a significant main effect of probe relatedness, where related probes were less accurate than unrelated probes ($F_1(1,18) = 29.5$, $p < .001$, $\eta^2_p = .621$; $F_2(1,81) = 49.9$, $p < .001$, $\eta^2_p = .381$). There was also a significant main effect of sentence ambiguity, where ambiguous sentence probes were less accurate than unambiguous probes ($F_1(1,18) = 4.54$ $p = .047$, $\eta^2_p = .201$; $F_2(1,81) = 4.82$, $p = .031$, $\eta^2_p = .056$). The interaction between sentence ambiguity and probe relatedness did not reach significance ($F_1(1,18) = 2.66$, $p = .120$, $\eta^2_p = .129$; $F_2(1,81) = 2.49$, $p = .118$, $\eta^2_p = .030$), although, on average, the ambiguity cost was larger for related probes than unrelated probes (3% vs. 0.6% respectively).

Ambiguous Inappropriate Probes. The ANOVA comparing accuracy of the ambiguous inappropriate probes with that of the ambiguous unrelated probes showed a main effect of unrelated probe type, where accuracy was lower for inappropriate than unrelated probes ($F_1(1,18) = 34.4$, $p < .001$, $\eta^2_p = .657$; $F_2(1,81) = 28.3$, $p < .001$, $\eta^2_p = .259$).

Reaction Times

The correct RT data were transformed and analysed using the same method as Experiment 1. Again, for ease of interpretation, the means of the untransformed data are presented (see Figure 2-2B middle column).

Sentence Ambiguity and Probe Relatedness. This ANOVA showed no significant main effect of sentence ambiguity, ($F_1(1,18) = 2.34$, $p = .114$, $\eta^2_p = .115$; $F_2(1,81) = 2.12$, $p = .150$, $\eta^2_p = .025$), although responses were on average, 23 ms slower for ambiguous than unambiguous probes. This difference was significant in the untransformed data ($F_1(1,18) = 5.34$, $p = .033$, $\eta^2_p = .229$; $F_2(1,81) = 4.49$, $p = .037$, $\eta^2_p = .053$). The main effect of relatedness was not significant in the transformed data ($F_1(1,18) = .062$, $p = .807$, $\eta^2_p = .003$; $F_2(1,81) = .725$, $p = .397$, $\eta^2_p = .009$) but was significant in the by-items analysis of the untransformed data, where responses, on average, slower for related than unrelated probes ($F_1(1,18) = .363$, $p = .555$, $\eta^2_p = .020$; $F_2(1,81) = 4.65$, $p = .034$, $\eta^2_p = .054$). There was no significant interaction between sentence ambiguity and probe relatedness ($F_1(1,18) = .012$, $p = .914$, $\eta^2_p = .001$; $F_2(1,81) = .214$, $p = .645$, $\eta^2_p = .003$).

Ambiguous Inappropriate Probes. Comparing the ambiguous-inappropriate and ambiguous-unrelated probe conditions, the results showed a significant main effect of unrelated probe type, where responses were 84 ms slower for the inappropriate probes ($F_1(1,18) = 13.4$, $p = .002$, $\eta^2_p = .427$; $F_2(1,81) = 27.6$, $p < .001$, $\eta^2_p = .254$).

2.3.3 Discussion

The pattern of accuracy across the different probe conditions was remarkably similar to that of Experiment 1 (see Figure 2-2A). Despite having an additional 200ms to process the sentence before the probe appeared, accuracy was still lowest for the ambiguous-related and the ambiguous-inappropriate probes. The ambiguous-inappropriate probes were, again, significantly less accurate than the ambiguous-unrelated probes, indicating a specific difficulty in suppressing contextually inappropriate meanings. Although the ambiguous-related probes also appeared to have particularly low accuracy, the interaction between sentence ambiguity and probe relatedness did not reach statistical significance. Despite this, there was still a significant main effect of ambiguity, suggesting that, even at this longer probe delay, participants still had more difficulty understanding ambiguous than unambiguous sentences. Together, these results show that these (relatively rare) comprehension failures are not eliminated by giving an additional 200ms before presenting the probe. This pattern may suggest that disambiguation failures may be an evitable consequence of semantic ambiguities. However, before drawing such a conclusion, it is important to investigate whether comprehension improves with an even longer probe-delay that gives participants even more uninterrupted processing time to disambiguate the ambiguous sentence successfully. Experiment 3 examines this hypothesis with a 1 second probe-delay.

The pattern of response times, which examine trials where comprehension was successful, however, was markedly different to that seen in Experiment 1. Although responses were, on average, slower for ambiguous than unambiguous sentence probes, this main ambiguity effect was not statistically reliable and was only around a third of the size found in Experiment 1's results (23 ms vs. 62 ms). Importantly, the interaction between ambiguity and probe-delay is significant, reported formally in Section 2.5. This suggests that, when probe words are presented at 300ms, the contextually appropriate meanings of ambiguous sentences are not significantly less integrated than those of unambiguous sentences. In contrast, the ambiguous-inappropriate probe condition still showed a specific, large (84 ms), and highly significant impairment compared to ambiguous-unrelated probes, suggesting that the inappropriate meanings are still more active than unrelated meanings. Taken together, these findings suggest a dissociation between the time course of integrating the appropriate meaning and that of suppressing the inappropriate meaning, such that inappropriate meanings are briefly maintained alongside the correct interpretation.

The finding that the 300ms probe delay produced well-integrated contextually appropriate meanings (or, at least, meanings that were not significantly less integrated than those of unambiguous sentences) suggests this process is relatively rapid and, hence, is efficiently enhanced by disambiguating information. The additional finding that this integration develops in spite of highly activated inappropriate meanings suggests that forming a new understanding of the sentence may not be substantially impaired by such inappropriate representations.

The finding that the inappropriate probes still incurred a significant cost with the 300ms delay indicates that meaning suppression during reinterpretation has two cognitive features. First, encountering contextual information that is inconsistent with a current interpretation is not sufficient for it to be quickly suppressed. This is directly consistent with the reordered access model of ambiguity resolution which claims that contextual information only acts to enhance the contextually appropriate meaning (Duffy et al., 2001; Duffy et al., 1988). It is also compatible with several other models of ambiguity resolution, including probabilistic constraint-based theories (MacDonald et al., 1994) and the structure-building framework (Gernsbacher, 1990; Gernsbacher & St John, 2001) that assume variable effects of context on the activation of an ambiguous word’s meanings. According to these models, contextual effects are dependent on its association with the meaning and the weighting of other information such as the meaning’s current level of activation. Second, the results demonstrate that activation of the appropriate meaning does not lead to simultaneous suppression of the inappropriate meaning. This is not predicted by models in which meanings are mutually exclusive such as Rodd et al.’s (2004) distributed connectionist model where increases in activation of the appropriate meaning must necessarily correspond to immediate decreases in activation of inappropriate meanings. However, this finding may be accommodated if there are multiple levels of semantic representations such that word meanings can be activated at an early lexical-semantic level of representation that is separate from the subsequent sentential representation in which only one meaning is currently integrated. Multiple types of representations are explicitly incorporated in the structure-building framework (Gernsbacher, 1990) and have been formally and successfully implemented in computational simulations of ambiguity resolution (Gernsbacher & St John, 2001).

Furthermore, the finding that the inappropriate meaning was still active when the probe was presented at 300ms contrasts with previous research on the initial selection of an ambiguous word’s meaning. Such research suggests that meanings may be selected within 200ms of hearing an ambiguous word (e.g., Seidenberg et al., 1982; Swinney, 1979). Thus, the question remains, how long does it take to suppress the inappropriate meaning during semantic reinterpretation? Studies on initial meaning selection that used the same semantic relatedness task as this current study used longer probe delays and report that good-readers can suppress inappropriate meanings by 750 ms or 1 second after reading an ambiguous word (Gernsbacher & Faust, 1991; Gernsbacher et al., 1990). Based on these studies, Experiment 3 presented the probe 1 second after the disambiguating word.

2.4 Experiment 3: 1 Second Probe Delay

2.4.1 Method

Participants

25 (14 female) participants from the same population as Experiment 1 took part in this experiment. Their mean age was 22 years (range: 18-34).

Stimuli, Design and Procedure

The procedure was the same as Experiment 1 except that the probe was presented at 1 second post sentence offset.

2.4.2 Results

Data Preparation and Analysis

Data were trimmed in the same way as Experiments 1 and 2 leading to one set of items (where the ambiguous word was “coat”) being removed from all subsequent statistical analyses. The same analyses were conducted as in Experiment 1 and 2. Individual responses faster than 300 ms or made after the timeout (2500 ms) were excluded, contributing to 0.6% of the total data.

Accuracy

Like Experiment 1 and 2, overall accuracy was high, above 90%, for all probe types with the lowest accuracy for ambiguous-related and ambiguous-inappropriate probes (see Figure 2-2A far right column).

Sentence Ambiguity and Probe Relatedness. This ANOVA showed a main effect of probe relatedness, where related probes were less accurate than unrelated probes ($F_1(1,18) = 18.2$, $p < .001$, $\eta^2_p = .502$; $F_2(1,83) = 21.17$, $p < .001$, $\eta^2_p = .203$). The main effect of sentence ambiguity was marginally significant in the by-subjects analysis, where ambiguous sentence probes were, on average, less accurate than unambiguous probes ($F_1(1,18) = 4.25$ $p = .054$, $\eta^2_p = .191$; $F_2(1,83) = 2.60$, $p = .111$, $\eta^2_p = .030$). The interaction between sentence ambiguity and probe relatedness was significant in the by-subjects analysis only, where the ambiguity cost was larger for related than unrelated probes ($F_1(1,18) = 7.30$, $p = .015$, $\eta^2_p = .288$; $F_2(1,83) = 2.81$, $p = .097$, $\eta^2_p = .033$). Additional ANOVAs conducted on related and unrelated probes separately revealed that the ambiguity effect was only significant for the related probes ($F_1(1,18) = 7.70$, $p = .012$, $\eta^2_p = .300$; $F_2(1,83) = 3.29$, $p = .073$, $\eta^2_p = .038$) but not for the unrelated probes ($F_1(1,18) = .122$, $p = .731$, $\eta^2_p = .007$; $F_2(1,83) = .049$, $p = .825$, $\eta^2_p = .001$).

Ambiguous Inappropriate Probes. The ANOVA comparing accuracy of the ambiguous inappropriate probes with that of the ambiguous unrelated probes showed a main effect of unrelated probe type, where accuracy was lower for inappropriate than unrelated probes ($F_1(1,18) = 28.3$, $p < .001$, $\eta^2_p = .611$; $F_2(1,83) = 24.4$, $p < .001$, $\eta^2_p = .227$).

Reaction Times

The correct RT data were transformed and analysed using the same method as the previous experiments. Again, for ease of interpretation, the means of the untransformed data are presented (see Figure 2-2B far right column).

Sentence Ambiguity and Probe Relatedness. This ANOVA showed a significant main effect of sentence ambiguity, with responses, on average, 48 ms slower for ambiguous than unambiguous probes ($F_1(1,18) = 27.2$, $p < .001$, $\eta^2_p = .602$; $F_2(1,83) = 17.0$, $p < .001$, $\eta^2_p = .170$). The main effect of relatedness was significant in the by-items analysis and marginally significant in the by-subjects analysis, with responses, on average, faster for related than unrelated probes ($F_1(1,18) = 4.12$, $p = .057$, $\eta^2_p = .186$; $F_2(1,83) = 17.6$, $p < .001$, $\eta^2_p = .175$). There was no significant interaction between sentence ambiguity and probe relatedness ($F_1(1,18) = 1.41$, $p = .250$, $\eta^2_p = .073$; $F_2(1,83) = .753$, $p = .388$, $\eta^2_p = .009$).

Ambiguous Inappropriate Probes. Comparing the ambiguous inappropriate and ambiguous unrelated probe conditions, the results showed no significant main effect of unrelated probe type with responses, on average, 16 ms faster for inappropriate probes than unrelated probes ($F_1(1,18) = 1.41$, $p = .250$, $\eta^2_p = .073$; $F_2(1,83) = 2.37$, $p = .127$, $\eta^2_p = .028$).

2.4.3 Discussion

The main aim of Experiment 3 was to assess the suppression of the inappropriate meaning further by using a longer probe delay, as both Experiment 1 and 2 revealed that this meaning was still more active than unrelated meanings with relatively short probe delays.

Accuracy

Although error rates were somewhat lower than in the previous experiments, the overall pattern of accuracy across the conditions remained similar. Although the ambiguity cost for ambiguous-related probes was only statistically reliable in the by-subjects analysis, this effect provides some evidence that, even with this longer probe delay, participants are still incurring disambiguation failures. More convincing evidence for these failures is seen in the performance on the ambiguous-inappropriate probes, where accuracy was significantly lower than for the control ambiguous-unrelated probes across both items and subjects, suggesting that inappropriate interpretations may be particularly pervasive even when participants are given additional time to process the sentence before encountering the visual probe.

Reaction Times

The pattern of response times across the conditions was markedly different to that seen in the earlier experiments. Fulfilling the aim of this experiment, presenting the probe 1 second after the disambiguating word eliminated the specific RT cost for the ambiguous-inappropriate probe condition seen in Experiment 1 and 2. In fact, the inappropriate probes showed a contrasting trend to be faster than ambiguous-unrelated probes, suggesting that, at this later time point, on those trials where disambiguation is achieved, the inappropriate meanings have been substantially suppressed compared to their activation at 300ms.

An unexpected finding was the re-emergence of a significant main effect of ambiguity. I suggest that this is not due to a re-emergent processing cost associated with accessing the appropriate meaning for two reasons. First, Experiment 2 showed only a minimal ambiguity cost, suggesting that, already when the probe is presented at 300 ms post sentence offset, appropriate meanings of ambiguous words are relatively well integrated. Second, unlike Experiment 1 and 2, with this 1 second delay the response times across all ambiguous probe conditions are remarkably similar to each other. This suggests there may be another processing cost that is affecting all three types of ambiguous probes. A possible explanation is that it is driven by meta-linguistic processes that are evoked more by the ambiguous sentences than the unambiguous sentences. For example, using ambiguous sentences that require reinterpretation may have made the ambiguity salient and the 1 second delay may have encouraged participants to reflect on the structure of those sentences and how the ambiguous words are resolved (i.e., by the last word). This may have generally delayed participants' responses to probes that followed ambiguous sentences. The semantic priming literature provides some indirect support for this hypothesis. Long delays of 1 second or more between semantically-related prime and target words (e.g., cat – dog) have been found to induce strategic or expectancy factors that affect lexical decisions and naming of the targets (e.g., Balota, Black, & Cheney, 1992; Hagoort, 1993; Neely, 1977). Neuroimaging research has additionally demonstrated that such long prime-target delays engage different neural correlates than short delays during lexical decision (Gold et al., 2006). Although these findings are based on tasks that implicitly tap into lexical-semantic processing, they provide evidence that long delays between semantic elements of a task may give participants time to process the stimuli in a qualitatively different way.

2.5 Experiments 1-3 Combined

The three experiments revealed several interesting effects of ambiguity. As some of these effects (particularly in the response times) appeared to be markedly different across the three probe delays, the three datasets were subjected to a combined analysis to examine the statistical reliability of these differences. To allow for direct comparisons across experiments, any sets of items which were excluded from at least one experiment's analyses were also excluded from the other datasets; this removed four sets of items. The analyses conducted were the same as those conducted for each experiment separately except that, now, probe delay (i.e., experiment) was included as a three-level (100ms, 300ms, 1 second) between-subjects factor in the by-subjects ANOVAs and as a within-items factor in the by-items ANOVAs.

Accuracy

Sentence Ambiguity, Probe Relatedness and Probe Delay. This ANOVA showed a main effect of probe relatedness, where related probes were less accurate than unrelated probes ($F_1(1,54) = 61.7$, $p < .001$, $\eta^2_p = .533$; $F_2(1,80) = 74.0$, $p < .001$, $\eta^2_p = .481$). The main effect of sentence ambiguity was also significant, where ambiguous sentence probes were less accurate than unambiguous probes ($F_1(1,54)$

= 15.5, $p < .001$, η^2_p= .223; $F_2(1,80)$ = 11.8, p = .001, η^2_p= .129). This main effect of ambiguity was modulated by a significant interaction between sentence ambiguity and probe relatedness in that the ambiguity cost was larger for related than unrelated probes ($F_1(1,54)$ = 12.0, p = .001, η^2_p=.182; $F_2(1,80)$ = 6.04, p = .016, η^2_p= .070). Additional ANOVAs conducted on related and unrelated probes separately revealed that the ambiguity cost was only significant for the related probes ($F_1(1,54)$ = 16.3, $p < .001$, η^2_p= .232; $F_2(1,80)$ = 9.86, p = .002, η^2_p= .110); there was no significant ambiguity effect for the unrelated probes ($F_1(1,54)$ = 1.91, p = .173, η^2_p= .034; $F_2(1,80)$ = 1.78, p = .186, η^2_p= .022).

Additionally, there was a main effect of probe delay that was significant only in the items analysis ($F_1(2, 54)$ = 2.52, p = .090, η^2_p= .085; $F_2(2,160)$ = 12.6, $p < .001$, η^2_p= .136). This main effect was largely driven by a probe delay by probe relatedness interaction that was again only significant in the items analysis ($F_1(2, 54)$ = 2.91, p = .063, η^2_p= .097; $F_2(2,160)$ = 7.34, p = .001, η^2_p= .084). Additional ANOVAs conducted on the related and unrelated probes separately revealed that the effect of probe delay was only significant (or marginal) for the related probes ($F_1(2,54)$ = 3.01, p = .057, η^2_p= .100; $F_2(2,160)$ = 11.7, $p < .001$, η^2_p= .128); there was no significant effect of delay for the unrelated probes ($F_1(1,54)$ = .388, p = .680, η^2_p= .014; $F_2(2,160)$ = 1.23, p = .294, η^2_p= .015). To statistically compare which of these accuracy differences were significant, additional ANOVAs were conducted on the related probe condition (averaging across ambiguity) for each pair of probe delay conditions. These analyses showed that accuracy for related probes was significantly higher with the 1 second probe delay than with either the 100 ms delay ($F_1(1,36)$ = 3.37, p = .075, η^2_p= .086; $F_2(1,80)$ = 10.4, p = .002, η^2_p= .115) or the 300 ms delays ($F_1(1,36) = 6.69$, $p = .014$, $\eta^2_p = .157$; $F_2(1,80) = 25.5$, $p < .001$, $\eta^2_p = .241$). There was no significant difference between the accuracy rates of the 100 ms and 300 ms delays (both Fs < 1).

Most importantly, there was no interaction between sentence ambiguity and probe delay nor a significant three-way interaction between ambiguity, relatedness and delay (all ps > .2), indicating that the impact of ambiguity did not differ significantly across the three experiments.

Ambiguous Inappropriate Probes and Probe Delay. The ANOVA comparing accuracy of the ambiguous-inappropriate probes with that of the ambiguous-unrelated probes showed a main effect of unrelated probe type, where accuracy was lower for inappropriate than unrelated probes ($F_1(1,54) = 89.6$, $p < .001$, $\eta^2_p = .624$; $F_2(1,80) = 50.3$, $p < .001$, $\eta^2_p = .386$). There was a main effect of probe delay but, again, this was only significant in the by-items analysis ($F_1(2,54) = 1.28$, $p = .286$, $\eta^2_p = .045$; $F_2(2,160) = 3.79$, $p = .025$, $\eta^2_p = .045$). ANOVAs comparing each pair of probe delays (averaging across probe condition), showed that accuracy was significantly higher with 1 second delay than with either the 100 ms delay ($F_1(1,36) = 1.84$, $p = .183$, $\eta^2_p = .049$; $F_2(1,80) = 4.30$, $p = .041$, $\eta^2_p = .051$) or the 300 ms delays ($F_1(1,36) = 2.25$, $p = .142$, $\eta^2_p = .059$; $F_2(1,80) = 6.63$, $p = .012$, $\eta^2_p = .077$). There was no reliable difference between 100 ms and 300 ms accuracy (both Fs < 1).

Again most importantly, the unrelated probe type by probe delay interaction was not significant by-subjects or by-items (ps > .2) indicating that the impact of ambiguity did not differ significantly across the three experiments.

Reaction Times

Sentence Ambiguity, Probe Relatedness and Probe delay. This ANOVA showed a significant main effect of sentence ambiguity, where responses were slower for ambiguous than unambiguous probes ($F_1(1,54) = 45.4$, $p < .001$, $\eta^2_p = .457$; $F_2(1,80) = 18.4$, $p < .001$, $\eta^2_p = .187$):. There was also a marginally significant main effect of relatedness, whereby responses were faster for related than unrelated probes ($F_1(1,54) = 4.00$, $p = .051$, $\eta^2_p = .069$; $F_2(1,80) = 3.29$, $p = .074$, $\eta^2_p = .039$). (The effect of relatedness was not significant in the untransformed data ($F_1(1,54) = .435$, $p = .512$; $F_2(1,80) = .080$, $p = .778$)).There was no significant sentence ambiguity by probe relatedness interaction (ps > .1).

There was a main effect of probe delay but it was significant only in the by-items analysis ($F_1 < 1$, $\eta^2_p = .035$; $F_2(2,160) = 90.2$, $p < .001$, $\eta^2_p = .530$). ANOVAs comparing each pair of probe delays (averaging across probe conditions) showed that responses were significantly faster with 100 ms delay than with either the 300 ms ($F_1(1,36) = 2.13$, $p = .153$, $\eta^2_p = .056$; $F_2(1,80) = 227.3$, $p < .001$, $\eta^2_p = .740$) or 1 second delay ($F_1 < 1$, $\eta^2_p = .008$; $F_2(1,80) = 19.1$, $p < .001$, $\eta^2_p = .193$). Responses were also faster with 1 second than 300 ms delay ($F_1 < 1$, $\eta^2_p = .019$; F_2 1,80) = 71.3, $p < .001$, $\eta^2_p = .471$). In the by-items analysis, there was also a significant interaction between probe delay and probe relatedness ($F_1(2,54) = 1.17$, $p = .319$, $\eta^2_p = .041$, $F_2(2,160) = 8.91$, $p < .001$, $\eta^2_p = .100$). ANOVAs analysing the effect of probe relatedness at each delay showed that the main effect of relatedness was significant in the 1 second delay condition ($F_1(1,18) = 4.40$, $p = .050$, $\eta^2_p = .196$; $F_2(1,80) = 16.0$,

$p < .001$, $\eta^2_p = .167$) but not with the 100 ms ($F_1(1,18) = 1.05$, $p = .320$, $\eta^2_p = .055$; $F_2(1,80) = 2.39$, $p = .126$, $\eta^2_p = .029$) or 300 ms delays (both Fs < 1).

Most importantly, probe delay significantly interacted with sentence ambiguity in both the by-subject and by-items analyses ($F_1(2,54) = 5.03$, $p = .010$, $\eta^2_p = .157$; $F_2(2,160) = 5.56$, $p = .005$, $\eta^2_p = .065$). (The sentence ambiguity by probe delay interaction was marginally significant in the untransformed data ($F_1(2,54) = 3.09$, $p = .054$; $F_2(2,160) = 2.90$, $p = .058$)). Additional ANOVAs were conducted to examine this interaction for each pair of probe delays. The ANOVAs revealed that the effect of sentence ambiguity was significantly smaller in the 300 ms condition (mean ambiguity effect = 23 ms) compared to both the 100 ms condition (59 ms: $F_1(1,36) = 8.18$, $p = .007$, $\eta^2_p = .185$; $F_2(1,80) = 9.77$, $p = .002$, $\eta^2_p = .109$) and the 1 second condition (47 ms: $F_1(1,36) = 6.56$, $p = .015$, $\eta^2_p = .154$; $F_2(1,80) = 5.01$, $p = .028$, $\eta^2_p = .059$). (The sentence ambiguity by probe delay interaction that compared the 1 second and 300 ms delays, approached significance in untransformed data ($F_1(1,36) = 3.48$, $p = .070$; $F_2(1,80) = 1.96$, $p = .165$)). The effect of sentence ambiguity was not significantly different between the 100 ms and 1 second conditions (both ps > .2).

As the probe delay condition (300 ms) that had the reduced ambiguity effect also had generally slower response times, it was important to assess whether reaction time was correlated with size of the ambiguity effect. To assess this, a Pearson’s moment correlation was conducted between the average RT and the average ambiguity cost (i.e., mean ambiguous probe RT – mean unambiguous probe RT) across the three probe delays. This showed a positive correlation between response time and ambiguity effect ($r_1(71) = .328$, $p = .005$; $r_2(85) = .359$, $p = .001$). That is, slower responses times correlated with *larger* ambiguity costs, and thus, slower RTs cannot explain the *smaller* ambiguity effect found in the 300 ms delay condition.

Moreover, the three-way interaction between sentence ambiguity, probe relatedness and probe delay was not significant by-subjects or by-items ($ps > .1$).

Ambiguous Inappropriate Probes and Probe Delay. Comparing the ambiguous inappropriate and ambiguous unrelated probe conditions, the results showed a significant main effect of unrelated probe type, with responses, on average, 36 ms slower for inappropriate than unrelated probes ($F_1(1,54) = 8.83$, $p = .004$, $\eta^2_p = .141$; $F_2(1,80) = 9.05$, $p = .004$, $\eta^2_p = .102$). There was also a main effect of probe delay that was significant in the items analysis only ($F_1 < 1$, $\eta^2_p = .020$; $F_2(2,160) = 19.1$, $p < .001$, $\eta^2_p = .193$). Responses were significantly slower with the 300 ms probe delay than with either the 100 ms delay ($F_1\ (1,36) = 1.00$, $p = .324$, $\eta^2_p = .027$; $F_2(1,80) = 27.9$, $p < .001$, $\eta^2_p = .259$) or 1 second delay ($F_1 < 1$, $\eta^2_p = .021$; $F_2(1,80) = 33.9$, $p < .001$, $\eta^2_p = .298$). There was no difference between 100 ms and 1 second delays (both $Fs < 1$).

Importantly, there was a significant interaction between unrelated probe type and probe delay ($F_1(2,54) = 7.68$, $p = .001$, $\eta^2_p = .221$; $F_2(2,160) = 13.7$, $p < .001$, $\eta^2_p = .146$). Additional ANOVAs conducted to assess this interaction between each pair of probe delays conditions showed that the inappropriate probe effect was significantly smaller with the 1 second probe delay than with both the 100 ms ($F_1(1,36) = 8.40$, $p = .006$, $\eta^2_p = .189$; $F_2(1,80) = 12.8$, $p = .001$, $\eta^2_p = .138$) and 300 ms probe delay ($F_1(1,36) = 12.1$, $p = .001$, $\eta^2_p = .251$; $F_2(1,80) = 24.4$, $p < .001$, $\eta^2_p = .234$). Although the inappropriate probe effect was larger in the 300 ms than the 100 ms delay, a further ANOVA that included only these two probe delays showed no significant interaction between probe delay and unrelated probe type, demonstrating that this difference was not statistically reliable ($F_1(1,36) = 1.01$, $p = .321$, $\eta^2_p = .027$; $F_2(1,80) = 3.02$, $p = .086$, $\eta^2_p = .036$). (The unrelated probe type by probe delay interaction that compared the 100 ms and 300 ms delays was significant in the items analysis on the untransformed data ($F_2(1,80) = 6.33$, $p = .014$)).

2.6 General Discussion

Understanding how ambiguous words are reinterpreted in light of disambiguating information can provide important insights into the flexibility of semantic interpretations. This study employed a new task to assess the effectiveness of reinterpretation as well as its time course. The results showed various ambiguity effects in both the accuracy and RT data which demonstrated that resolving late-disambiguation sentences is a cognitively costly but fully achievable process.

Effectiveness of Semantic Reinterpretation

The accuracy data showed interesting effects of ambiguity that were not significantly modulated by the length of the probe delay. First, across the three experiments, accuracy was high. Participants were, on average, over 90% accurate for the two critical ambiguous conditions: ambiguous-related and ambiguous-inappropriate probes. This indicates that listeners are usually highly effective at reinterpreting dominant-meaning interpretations of ambiguous words in light of inconsistent, subordinate-biased, information. At first, this finding may not seem surprising because the sentences contained disambiguating words. However, several features of the disambiguating context may have made it difficult for participants to resolve these ambiguities correctly. The disambiguating context was presented at the end of a fairly long sentence, several words after the ambiguous word. Such distance could have impaired integration of these sentential elements due to memory constraints. This may have been particularly detrimental as the sentences were spoken, and, thus, do not leave a permanent signal that may be revisited like written sentences. In addition, the context was minimal in that only one, or occasionally two, words were used, which could have been inadequate to trigger reinterpretation of an inconsistent interpretation. Instead, the high accuracy for both ambiguous-related and inappropriate probes shows that listeners make use of even minimal context to resolve ambiguity and, thus, must be constantly integrating contextual information to ensure accurate comprehension of words heard much earlier in sentences. This concurs with probabilistic constraint-based theories of language comprehension, in which multiple sources of information are continually and simultaneously used to constrain comprehension (MacDonald et al., 1994; MacDonald & Seidenberg, 2006). In addition, these findings for spoken sentences complement and converge with the high accuracy found for visually-presented late-disambiguation sentences (Miyake et al., 1994; Zempleni et al., 2007), despite the potentially different processing demands that may result from the inherently different signals (i.e., transient versus permanent).

Despite high absolute accuracy, accuracy was reduced for the ambiguous-related and ambiguous-inappropriate probes (relative to the appropriate control conditions). Because performance in these conditions did not significantly improve with increasing probe delay and there was still evidence of occasional disambiguation errors with the 1 second probe delay, this suggests that a small number of misinterpretations might be inevitable wherever semantic ambiguity of this type is present. This failure to appreciate the relationship between context and the intended meaning of the ambiguous word may be the result of numerous factors, such as (i) lapses in attention toward the disambiguating information, (ii) idiosyncratic associations that listeners hold between the disambiguating words and the *un*intended meaning, or (iii) for some listeners, the intended meaning may be too infrequent to be activated by the context (Rodd et al., 2013). It must be noted, however, that these errors only occurred on a small portion of trials (less than 10%), demonstrating that listeners are usually highly effective at resolving ambiguities. Future work is needed to assess the causes of disambiguation failures and possible individual differences in the efficiency of these processes (Gernsbacher et al., 1990; Twilley & Dixon, 2000).

Time Course of Semantic Reinterpretation

Probe delay significantly modulated two ambiguity effects in the RT data: the main effect of ambiguity and the inappropriate probe effect. These modulations provide insights into the time course of the processes underlying semantic reinterpretation. Experiment 1 revealed a significant main effect of sentence ambiguity when probes were presented 100 ms after the disambiguating word. Both ambiguous-related and ambiguous-unrelated probes were responded to more slowly than unambiguous sentence probes, suggesting that disambiguation was not complete by the time the probe word appeared. This suggests that semantic reinterpretation is not a process that occurs rapidly when inconsistent information is encountered and, hence, is consistent with the notion that reinterpretation is a time consuming, cognitively demanding process (Duffy et al., 1988; Rayner & Frazier, 1989; Rodd, Johnsrude, et al., 2010; Zempleni et al., 2007).

Importantly, there was a significant interaction between this general ambiguity effect and probe delay. Critically, the ambiguity effect significantly reduced to around a third of its size when the probe was presented 300 ms after the disambiguating word. Although there was some evidence that responses were also generally slower with the 300 ms probe delay compared to the 100 ms delay, this cannot explain the reduced ambiguity cost found with this delay because 1) the slow RT trend was only reliable across items, suggesting it was driven by a subset of participants and 2) slower RTs significantly correlated with *larger* ambiguity costs not smaller ones. Thus, instead, these findings suggest that having an extra 200 ms of uninterrupted processing time helped enhance the meaning of the sentence. This finding, thus, further suggests that the processing cost of reanalysing ambiguity is a relatively transient phenomenon reflecting relatively rapid reinterpretation of the correct meaning. Unexpectedly, this ambiguity cost re-appeared with the longest probe delay (1 second). However, this may be driven by meta-linguistic awareness of the ambiguity that is evoked by long delays. This issue will be discussed further later.

Interestingly, there was no interaction between sentence ambiguity and probe relatedness in any of the three experiments, indicating that there was no specific slow down for probes that were related (as opposed to unrelated) to the ambiguous words. This is most likely because the task required explicit semantic decisions, meaning that all probes rely on participants having integrated the correct meaning. Thus, partial activation of, or uncertainty about, the contextually appropriate meaning can slow down decisions for both related and unrelated probes.

The second important ambiguity effect was a specific RT cost for ambiguous-inappropriate probes. In contrast to the main effect of ambiguity, this inappropriate probe effect was evident with both the 100 ms and 300 ms probe delays. The finding that RTs for these probes were greater than the ambiguity-unrelated probes indicates that this processing cost extended beyond the general cost of ambiguity and, therefore, can specifically suggest that inappropriate meanings were more active than unrelated meanings at these two time points. Unlike the general ambiguity cost that was significantly smaller at 300 ms compared to at 100 ms, the inappropriate cost was numerically larger (albeit not significantly) at 300 ms, suggesting that no substantial suppression of the inappropriate meaning occurred during this interval. Importantly, as the reduction in the general ambiguity cost with 300 ms probe delay was not accompanied by significant suppression of the inappropriate meaning, this supports the view that the general ambiguity cost is specifically associated with stability of the contextually appropriate meaning. Furthermore, the inappropriate probe cost disappeared with the 1 second probe delay, suggesting that the inappropriate meaning requires several hundred milliseconds to be reduced.

Theoretical Implications

Together, these error rate and response time data suggest that, when disambiguating information was encountered several words after an ambiguous word, contextually inappropriate meanings were over-active whilst contextually appropriate meanings were under-active. Time-costly processes, then, ensued to suppress the former and enhance the latter. This is consistent with a wealth of behavioural research and many models of semantic ambiguity resolution which argue that listeners (and readers) rapidly select a single interpretation of an ambiguous word even when no disambiguating context has yet been encountered, which will require reinterpretation if it is not consistent with subsequent information (Duffy et al., 2001; Duffy et al., 1988; Rodd, Johnsrude, et al., 2010; Twilley & Dixon, 2000).

Current perspectives on ambiguity resolution, however, lack detail about the processes underlying semantic reinterpretation, stating only two features: 1) that it is a time-consuming or cognitively demanding operation and 2) that it involves two key processes of selecting the appropriate meaning and suppressing the inappropriate meaning (Duffy et al., 1988; Rodd, Johnsrude, et al., 2010, 2012). The current results provide novel evidence that support and detail these features further. First, the main ambiguity effect found in the RT data constitutes the first direct evidence that integrating a non-selected, yet, contextually appropriate meaning is a time-consuming process, as speculated in previous studies that found general processing costs associated with reinterpretation (e.g., Duffy et al., 1988; Rayner & Duffy, 1986; Rodd, Johnsrude, et al., 2010). This process may be slowed through interference from contextually inappropriate meanings that were initially selected during comprehension, as evinced by the inappropriate probe RT cost. However, the additional finding that the general ambiguity effect reduces without a corresponding reduction in the inappropriate probe cost suggests that the activation of an inappropriate meaning does not prevent listeners from ultimately attaining a correct interpretation of the sentence. This, perhaps, suggests a dissociation between meanings being active and their being integrated into a listener's current interpretation of a sentence (or discourse), which is conceivably akin to Gernsbacher and St John's (2002) distinction between concept-level and sentence-level representations.

Furthermore, the results provide the first evidence suggesting that the two key processes of activating appropriate meanings and suppressing inappropriate meanings during reinterpretation have different time courses. Unfortunately the exact timings of these processing stages cannot be inferred on the basis of this task because a decision is not made instantaneously with presentation of the probe. Instead, the different patterns of responses at these different probe positions provide insights into the relative timing of these two critical processes. The results suggest that the former is a relatively rapid process while the latter suppression process has a slower time course. This is difficult to reconcile with models in which meanings are strictly mutually exclusive such as distributed connectionist models (e.g., Rodd et al., 2004) unless multiple levels of representation are assumed (Gernsbacher & St John, 2001).

Furthermore, while the current study cannot explain the reasons for this dissociation, the notion that a dissociation may occur is consistent with various different theories of ambiguity resolution. The reordered access model suggests that contextual selectivity may drive this difference, since contextual information is assumed to selectively affect the contextually appropriate meaning, leaving the inappropriate meanings to decay passively (Duffy et al., 2001; Duffy et al., 1988). In contrast, MacDonald et al.'s (1994) constraint-based theory does not claim for exclusive effects of specific constraints on specific meanings but suggests this dissociation may result from different weightings of constraints for alternative meanings. Under this account, the reinterpretation dissociation may arise for two reasons. First, the contextual information may have supported the contextually appropriate meaning more than it constrained the inappropriate meaning. This is possible as the disambiguating words in this set of stimuli explicitly related to features of the contextually appropriate meaning. For example, in the sentence "the teacher explained that the bark was going to be *damp*", the word "damp" relates to the texture of the tree meaning of "bark". Second, the activation levels of the two meanings may have differently modulated the influence of the contextual constraints. For instance, inappropriate meanings that were initially selected may have developed robust activation during sentence comprehension or been maintained by some excitatory connections, which may render these meanings difficult to suppress. Gernsbacher et al.'s structure building framework (Gernsbacher, 1990; Gernsbacher & St John, 2001) posits an additional more temporally specific explanation. According to this theory, suppression of inappropriate meanings may be delayed because it is driven, and thus must be preceded, by the creation of the new sentence-level representation. An alternative entirely different explanation is that currently non-preferred meanings may be actively maintained (e.g., by executive control mechanism) in case the current meaning turns out to be incorrect, as argued by right-hemisphere accounts of ambiguity resolution (e.g., Faust & Chiarello, 1998; Faust & Gernsbacher, 1996).

Despite the apparent temporal dissociation between the activation and suppression of meanings, the results showed that, in the 1 second probe delay condition, the inappropriate meaning was reduced to the level of unrelated word meanings. This is entirely compatible with most ambiguity models which view disambiguation as the process of settling on one interpretation whilst actively suppressing/inhibiting alternative meanings (Gernsbacher, 1990; MacDonald et al., 1994; Rodd et al., 2004) or letting them decay to their original state (Duffy et al., 2001; Duffy et al., 1988).

Methodological Implications

The results also highlight several methodological issues that are important to consider when investigating ambiguity resolution. The finding of a general processing cost of disambiguation calls into question the suitability of baseline conditions used in previous ambiguity studies. For example, Gernsbacher et al. (1990) pioneered the use of semantic relatedness tasks to assess activation of contextually inappropriate meanings, arguing that by using probes unrelated to unambiguous sentences as the comparison condition, any additional time required to respond to ambiguous-inappropriate probes would index how activated the inappropriate meaning was. However, the results of this study show that relatedness judgements on an ambiguous sentence can be delayed even when the decision (ambiguous-unrelated) does not pertain to the inappropriate meaning. Thus, slower response times for ambiguous-inappropriate probes compared to unambiguous-unrelated probes may reflect a general processing difficulty for ambiguous sentences rather than an activated inappropriate meaning per se. Hence, I argue for the use of an ambiguous sentence control condition (i.e., ambiguous-unrelated) to account for general costs of processing ambiguous sentences.

In addition, the unexpected re-emergence of the general ambiguity effect with the 1 second probe delay highlights the potential problem that long probe delays can cause by evoking processes not involved in natural sentence comprehension. The 1 second delay may have encouraged participants to reflect on the structure of those sentences causing participants to be generally delayed when presented with a probe after an ambiguous sentence. This possibility is supported by the finding that all ambiguous probe types had similarly long response times, suggesting that, with this longer delay, there is a factor (i.e. metal-linguistic awareness) affecting decisions on ambiguous probes in general. This further emphasizes the importance of ambiguous control conditions so as to assess general behavioural effects of ambiguity that are not specific to the alternative meanings of ambiguous words as well as including intermediate probe delays. It must be noted, however, that the use of these baselines may be more important for certain types of tasks and stimuli than others, in particular, for sentences or tasks likely to evoke meta-linguistic awareness of the ambiguity (e.g., late-disambiguation sentences and explicit semantic tasks).

A potential concern of using a probe task to assess comprehension is that the probe itself may affect the natural course of the comprehension process. In particular, the inappropriate interference effect that was found with the 300 ms, and even the 100 ms, probe delays may have occurred because the probe itself “reactivated” the inappropriate meaning of the ambiguous word rather than because the inappropriate meaning was still active when the probe appeared. While this is a possibility and could potentially contribute, to some degree, to the inappropriate interference effect, it cannot explain the full pattern of results. First, if the inappropriate probe (e.g., “bread”) is simply “reactivating” the inappropriate meaning of an ambiguous word (e.g., “toast” in “the man made the toast with a very old microphone”), then an interference effect would most likely have been found regardless of probe delay. However, when the probe was presented at 1 second post sentence offset, responses were not significantly slower for the inappropriate probes than the control ambiguous-unrelated probes. In fact, they were numerically faster. This suggests that a simple reactivation hypothesis is not adequate to explain the lack of an inappropriate interference effect at the longest probe delay. A more complex reactivation hypothesis may still be possible, where the level of interference from the "reactivated" meaning relies on the level of uncertainty that surrounds the sentence meaning. That is, reactivation of inappropriate meaning may not have interfered with comprehension at 1 second because the sentence meaning had developed into a stable state by that time such that it was impervious to influences from the inappropriate meaning. However, this does not explain the lack of difference between the inappropriate effects with the 100 ms and 300 ms delays, given that the results also showed a greater stability/certainty of the sentence meaning between these two time-points. This improvement should have then also reduced the level of interference caused by the inappropriate probe.

Thus, together the set of results suggests that the interference effect of the probe condition is not merely due to reactivating a non-active inappropriate meaning. The inappropriate meaning must be at some level of activation when the 100 ms and 300 ms probes were encountered in order to be susceptible to influence from the probe.

In addition, there is no behavioural evidence to suggest that this level of activation is qualitatively different between these time points. However, their activation does qualitatively change when the probe is presented at 1 second. Nevertheless, future studies are needed to test this reactivation hypothesis by using a more implicit semantic task that reduces the potential task-induced interference of the inappropriate probe, such as a word-naming task to assess whether the pattern of effects found with the probe verification task remain the same. Performance on these tasks has been shown to be relatively immune to backward priming effects of the probe word, as evidenced by the finding that targets that have strong backward associations with the prime (e.g., pan-bed) do not produce priming (e.g., Burgess, Tanenhaus, & Seidenberg, 1989; Seidenberg, Waters, Sanders, & Langer, 1984).

Future research

As well as providing novel findings to expand models of semantic disambiguation, this study also highlights various questions that warrant future research. First, to what extent is the suggested temporal dissociation between contextually appropriate and inappropriate meaning modulated by meaning frequency? This is important as frequency is consistently highlighted as an important constraint in ambiguity resolution (e.g., Chen & Boland, 2008; Duffy et al., 2001; Duffy et al., 1988; Gernsbacher & St John, 2001; MacDonald et al., 1994; Simpson, 1994; Tabossi & Zardon, 1993; Twilley & Dixon, 2000). Second, to what degree does the amount of disambiguating context (e.g., single vs. multiple biasing words) and/or the nature of this context (e.g., syntactic, pragmatic or semantic cues) affect the speed and quality of disambiguation? By assessing both the outcome of the disambiguation process via accuracy and the process of disambiguation via reaction times, this paradigm constitutes a novel efficient way to investigate these questions.

Furthermore, the reinterpretation process has, until now, been discussed in rather broad terms as a process that has two aspects: 1) integrating a non-preferred, yet, contextually appropriate meaning and 2) inhibiting a currently active, but inappropriate, meaning that was selected prior to the disambiguating information. However, these two components are unlikely to involve the same operations across all types of ambiguous sentences. In particular, their nature is inevitably dependent on the nature and outcome of the initial selection process. For example, for some sentences, selecting the contextually appropriate interpretation may involve activating a currently “inactive” meaning. This may be because that meaning was either suppressed earlier during comprehension or it was never activated at the initial encounter of the ambiguous word. The latter may be the case for words that have really infrequent meanings (e.g., the “animal enclosure” meaning of “pen”). For other sentences, this process may involve merely boosting the activation of meaning that has been maintained at a lower level of activation during comprehension. Future research is needed to distinguish between these possibilities.

Moreover, some researchers have also argued that meanings may be maintained in parallel until contextual support is encountered, especially if working memory capacity is sufficient for this (Miyake et al., 1994). Therefore, an additional possibility is that for some sentences listeners have maintained multiple meanings in parallel and the processing cost reflects difficulty in selecting between simultaneously active meanings as oppose to difficulty in reinterpretation per se. It is possible that such selection would be demanding in these kinds of sentences because of the long distance between the ambiguous word and the contextual information that may make contextual integration difficult or it may be demanding merely because both representations have been strengthened throughout this delay, rendering it difficult to select one representation. However, it is unlikely that late-selection is substantially contributing to the processing costs found in this study because the majority of sentences were biased in that the ambiguous words had one more dominant meaning. It is well-established from behavioural data, and shown in computational simulations, that frequency information guides activation and integration with the sentence context such that the more dominant meaning is maintained at a higher level and integrated more than the subordinate meaning (e.g., Duffy et al., 2001; Gernsbacher & St John, 2001; Simpson & Krueger, 1991; Twilley & Dixon, 2000), which will need reinterpretation when context supporting the subordinate meaning is encountered. Nevertheless it is possible that selection may be delayed for balanced ambiguous words which have two relatively equal frequencies (Miyake et al., 1994). On the other hand, there is a larger body of evidence from priming and eye-movement studies that converges on the conclusion that even for balanced words a single meaning is predominately integrated into comprehension (e.g., Duffy et al., 1988; Rayner & Duffy, 1986; Seidenberg et al., 1982; Swinney, 1979). Thus, it is clear that much more research is needed not only to understand the operations underlying reinterpretation but also the nature of the initial selection process.

Conclusion

The accuracy and reaction time results show that delaying disambiguating information until the end of a spoken sentence does not prevent successful and complete resolution of semantic ambiguity. Semantic representations generated during comprehension can be flexibly changed in light of conflicting context. However, delaying context causes listeners to select an initial interpretation prematurely, which requires costly reinterpretation when a contextually inappropriate meaning is selected. In addition, the results suggest that suppression of the contextually inappropriate meanings does not occur immediately upon encountering inconsistent disambiguating information but may be delayed for several hundred milliseconds. Furthermore, despite high rates of disambiguation, some miscomprehensions seem inevitable wherever ambiguities of this type are present.

Chapter 3 Neural Responses to Semantically Ambiguous Sentences: An fMRI Study

3.1 Introduction

A large body of behavioural research, including my work presented in the preceding chapter, demonstrates that understanding the correct meaning of an ambiguous word is a cognitively demanding process, especially when semantic reinterpretation is required. Existing neuroimaging studies of semantic ambiguity resolution have identified several key regions as being involved in this process (Bekinschtein, Davis, Rodd, & Owen, 2011; Davis et al., 2007; Mason & Just, 2007; Rodd et al., 2005; Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007, see Figure 3-1) but their precise roles remain uncertain. Examining the neural responses to semantic ambiguities is important as it provides a means to gain insight into high-level language networks operating during sentence comprehension that support key cognitive operations such as conflict resolution, contextual integration and inhibition.

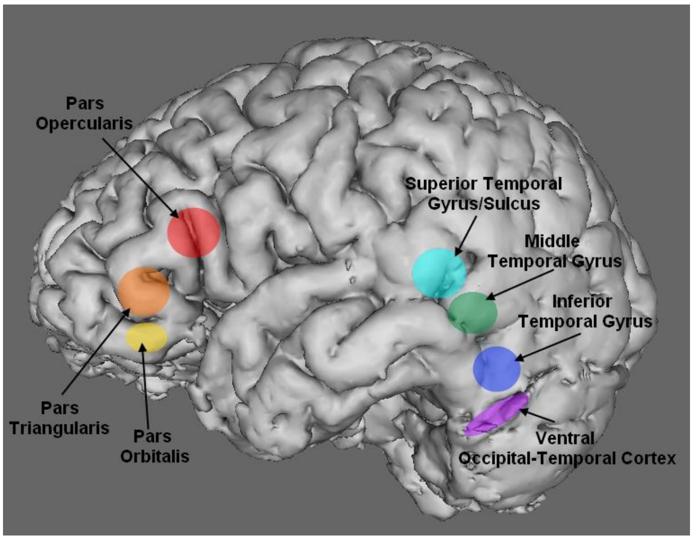

Figure 3-1: Left hemisphere regions that have been shown to have greater BOLD signal for semantically ambiguous than unambiguous sentences.

3.1.1 Initial Neuroimaging Studies

One of the first fMRI studies to investigate semantic ambiguity resolution within sentence contexts was conducted by Rodd et al. (2005). Participants listened to sentences that contained several ambiguous words (e.g., “there were *dates* and *pears* on the kitchen table”). Activation was greater for ambiguous than well-matched unambiguous sentences in several frontal and temporal regions: the left and right inferior frontal gyrus (LIFG, RIFG) and the left posterior temporal cortex that included the inferior temporal gyrus (LITG), fusiform and middle temporal gyrus (LMTG). These results provided the first evidence that the comprehension of semantically ambiguous sentences may be supported by a fronto-temporal network.

The specific regions that were engaged by ambiguity were theoretically important on several levels. First, the LIFG activation resided in its middle and posterior sub-divisions (pars triangularis and pars opercularis), which have been traditionally attributed to non-semantic processes, including speech production (Broca, 1861; Geschwind, 1970; Indefrey & Levelt, 2004; Penfield & Roberts, 1959; Wernicke, 1874) and syntactic computations (Caplan, Alpert, & Waters, 1998; Caramazza & Zurif, 1976; Dapretto & Bookheimer, 1999; Goodglass et al., 1979; Grodzinsky, 1986; Schwartz, Saffran, & Marin, 1980). Yet, Rodd et al.’s (2005) study tapped into semantic processing demands. Second, when prior research had attributed the LIFG to semantics, it was often highlighted as being important for explicit semantic decisions that required the strategic retrieval, maintenance or selection of semantic information rather than for natural speech comprehension per se (Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Gabrieli, Poldrack, & Desmond, 1998; Novick, Trueswell, & Thompson-Schill, 2005; Thompson-Schill, D'Esposito, Aguirre,

& Farah, 1997; A. D. Wagner, Paré-Blagoev, Clark, & Poldrack, 2001). However, in Rodd et al.'s (2005) study, LIFG activation was higher for ambiguous sentences despite the fact that participants were not aware of the ambiguity and did not rate these sentences as being any less natural than the unambiguous sentences. Such activation in the absence of explicit awareness makes it unlikely that they were due to strategic processes, and, hence, suggested a more routine involvement of this region in natural sentence comprehension. Rodd et al.'s (2005) findings are not, however, incompatible with non-semantic or strategy-based semantic accounts of LIFG function, as regions may serve multiple functions or there may be an overlap between neuronal populations that support different functions within the same anatomical regions (see Price & Devlin, 2003 for similar arguments about the left mid-fusiform gyrus). Furthermore, the posterior temporal activation was intriguing because it contrasted with accounts that highlighted anterior temporal structures as being important for semantic processing (Chan et al., 2001; Mummery et al., 2000; Nobre, Allison, & McCarthy, 1994). Thus, together, Rodd et al.'s (2005) study suggested that the LIFG may be more important for normal speech comprehension than had previously been emphasised, and that posterior, as oppose to anterior, temporal cortex is involved in semantic processes associated with ambiguous words.

The consistency of this network in ambiguity resolution was supported by two subsequent studies that used the same stimuli as Rodd et al. (2005). Davis et al. (2007) replicated these findings in a study that examined ambiguity resolution at different levels of awareness. Subjects who were fully awake, but not those who were lightly or deeply sedated, showed ambiguity-elevated activity in the LIFG bilaterally and in the posterior LITG. Such responses were also found in a small number of patients who were clinically diagnosed as being in a vegetative state or minimally conscious (Coleman et al., 2009; Coleman et al., 2007). Together, these results demonstrated that semantic ambiguity produces significant and reliable effects on neural activity, highlighting the importance of understanding this brain network for developing comprehensive models of language comprehension.

From these ambiguity studies it is unclear which ambiguity-related processes are supported by these regions. As reviewed in Chapter 1, the psycholinguistic literature emphasises two key cognitive components of semantic ambiguity resolution during sentence comprehension: 1) initial meaning selection, when a single meaning is quickly selected after an ambiguous word is encountered and 2) semantic reinterpretation, when the initial selection needs to be changed in light of subsequent, inconsistent, context. Such research also demonstrates that different types of ambiguous words and sentential contexts place differential demands on these two components (Duffy et al., 1988; Rayner & Duffy, 1986; Rodd, Johnsrude, et al., 2010; Sereno et al., 2006). However, the set of ambiguous sentences examined in the aforementioned fMRI studies were heterogeneous in terms of these properties. For example, the position of the ambiguous words and the disambiguating information varied across sentences, the ambiguous words were a mixture of balanced and biased words and the contextual bias was not explicitly manipulated. Thus, it is unclear which ambiguity demands were contributing to the activation of these regions.

Building on the initial work of Rodd et al. (2005), several fMRI studies have examined this network by systematically manipulating demands on different ambiguity processes (Bekinschtein et al., 2011; Mason & Just, 2007; Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007). Such research has revealed various important insights into the ambiguity network, especially regarding the potential roles of the LIFG and the left posterior temporal cortex in semantic ambiguity resolution.

3.1.2 LIFG and Semantic Ambiguity Resolution

The LIFG is the most consistent region to show increased BOLD signal for sentences that contain semantically ambiguous words. In fact, this region has been reported in all published studies directly comparing these sentences with unambiguous ones (Bekinschtein et al., 2011; Davis et al., 2007; Mason & Just, 2007; Rodd et al., 2005; Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007). Such consistency is due to the fact that the LIFG responds to a wide variety of semantically ambiguous sentences. Compared to unambiguous sentences, increased activation has been found for late-disambiguation sentences in which the ambiguous word was disambiguated to its subordinate meaning (Mason & Just, 2007; Rodd, Johnsrude, et al., 2012; Zempleni et al., 2007), dominant meaning (Zempleni et al., 2007) or one of two equally frequent (i.e., balanced) meanings (Mason & Just, 2007). In addition, such higher activation is also reported across a range of disambiguating positions: for sentences in which the disambiguating information immediately follows the ambiguous word (Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010), is delayed for several words (Mason & Just, 2007; Rodd, Johnsrude, et al., 2012; Zempleni et al., 2007), or even precedes the ambiguity (Rodd, Johnsrude, et al., 2012). Together, these findings clearly demonstrate that the LIFG plays a fundamental role in ambiguity resolution.

Involvement of the LIFG across a variety of ambiguous sentences suggests that this region may support the initial processing of an ambiguous word. Direct evidence for this role was found by Rodd et al. (2012) who used semisparse fMRI to examine the time-course of neural responses to specific ambiguity-related events during sentence comprehension. The results showed that LIFG activation for ambiguous sentences was greater than unambiguous sentences at time-points that corresponded to presentation of the ambiguous word and was evident for sentences in which the disambiguating information preceded the ambiguous word (e.g., “the *hunter* thought that the hare in the field was actually a rabbit”) as well as for sentences in which the context occurred much later than the ambiguity (e.g., “the ecologist thought that the plant by the river should be *closed down*). Based on these results, Rodd et al. (2012) argued that the LIFG supports initial semantic selection processes that occur when an ambiguous word is encountered, supporting the selection of a single meaning based on contextual or frequency information, as proposed in current psycholinguistic models of ambiguity resolution (e.g., Duffy et al., 2001; Twilley & Dixon, 2000). The importance of the LIFG in the initial processing of an ambiguous word during sentence comprehension is further corroborated by several patient studies. Such studies demonstrate that patients with Broca’s aphasia are impaired in using prior sentence context to activate or select the contextually appropriate meaning of ambiguous words (e.g., Grindrod, 2012; Swaab, Brown, & Hagoort, 1998; Swinney et al., 1989). Supporting this finding, Vuong and Martin (2011) recently showed that patients with relatively circumscribed lesions to the LIFG also had difficulty reading such sentences, especially when the context preceding the ambiguous word biased its subordinate meaning.

While such evidence clearly demonstrates that the LIFG has a role in the initial processing of an ambiguous word, the aforementioned fMRI studies also suggest that this region is involved further downstream during comprehension, specifically, when the meaning of an ambiguous word needs to be reinterpreted. Supporting evidence for the LIFG's role in semantic reinterpretation comes from the finding that its activation is affected by two factors: the position of the disambiguating information in the sentence and meaning dominance. As discussed in Chapter 2, reinterpretation is most likely to occur for late-disambiguation sentences, where disambiguating information is delayed for several words after the ambiguous word, since a meaning is selected relatively quickly after encountering an ambiguous word without prior context (Duffy et al., 1988; Rodd, Johnsrude, et al., 2010; Seidenberg et al., 1982; Swinney, 1979; Twilley & Dixon, 2000). Reinterpretation is especially likely for two variants of late-disambiguation sentences: 1) for those with a biased ambiguous word that is disambiguated to its subordinate meaning since the dominant meaning is likely to be initially selected (Duffy et al., 2001; Duffy et al., 1988; Simpson & Krueger, 1991; Twilley & Dixon, 2000) and 2) for those with a balanced ambiguous word. For the latter, the incorrect meaning is likely to be selected approximately half the time, as they do not have a systematic frequency bias (Duffy et al., 1988; Seidenberg et al., 1982; Swinney, 1979).

Three previous fMRI studies support the semantic reinterpretation hypothesis of the LIFG, showing increased activation in this region to late-disambiguation sentences and that it is affected by the position of disambiguation and modulated by meaning dominance. First, Rodd et al. (2012)'s time course study demonstrated that the LIFG produced responses that were time-locked to the disambiguating information as well as to the ambiguous word itself. Specifically, sentences in which disambiguating information was presented a few words after the ambiguous word (e.g., "the scientist thought that the <u>film</u> on the *water* was from the pollution") produced greater responses in early processing time-windows than sentences in which context was delayed until several words later ("e.g., the ecologist thought that the <u>plant</u> by the river should be *closed down*"). The latter produced more activation in later time-windows compared to the former. In support, Zempleni et al. (2007) showed greater LIFG activation for late-disambiguation sentences that referred to the subordinate meaning of biased ambiguous words than sentences that referred to the dominant meaning. In both types of sentences, the dominant meaning would often be initially selected, since context was presented later in the sentence. Thus, in the former sentences, this selection would turn out to be incorrect when the subordinate-biasing context would be encountered, and thus need reinterpretation, whereas it would be correct in the latter sentences and thus not need reinterpretation. Corroborating evidence was further reported by Mason and Just (2007) who compared late-disambiguation sentences that contained balanced words with those that contained biased words. The latter were always disambiguated toward the subordinate meaning. Initial support came from the finding that both types of sentences produced greater LIFG activation than unambiguous sentences. This is in line with the reinterpretation hypothesis, since, again, delaying context causes inappropriate meanings to be selected on some proportion of trials. However, this is not strong evidence since both contrasts compare ambiguous sentences to unambiguous sentences so the results may merely reflect processes associated with the initial processing of an ambiguous word rather than its reinterpretation per se. More critical support came from their direct comparison between the two types of ambiguous sentences, which showed that activation was additionally greater for the biased than balanced sentences. The former requires reinterpretation more often because it is only for these words that frequency systematically biases the initial meaning selection towards the incorrect (i.e., dominant) meaning. Therefore, together, this set of findings is consistent with the main predictions of the semantic reinterpretation hypothesis.

Supporting evidence, however, has not been found in all studies that examined late-disambiguation studies. Rodd et al. (2010) failed to find significant effects of meaning dominance on LIFG activation and Bekinschtein et al. (2011) only found a significant ambiguity response in this region for humorous, but not for non-humorous sentences. However, the lack of support for reinterpretation in these two studies can be explained in ways that do not have negative implications for this hypothesis. Rodd et al. (2010) suggested that the effect of meaning dominance may have been reduced in their study because of the relative position of the ambiguity and the disambiguating information in their stimuli. In their sentences, the ambiguous word was always disambiguated immediately after the ambiguity by the subsequent verb, which may have left insufficient time for a meaning to be selected prior to the disambiguating information that would need reinterpretation (e.g., “In Australia, strange <u>barks</u> *grow* on many trees”). Bekinschtein et al.’s (2011) null-finding for non-humorous sentences is more surprising as the ambiguous stimuli followed a similar structure to that used in previous studies where the disambiguating information was delayed relative to the ambiguous word. However, the ambiguous word and disambiguating information were not presented in the same sentence. Instead, the ambiguous word was presented at the end of a question and disambiguated by the following sentence (e.g., What happened to the <u>post</u>? As usual, it was given to the best-qualified *applicant*). This may have caused the ambiguity and its reinterpretation to be processed differently compared to when an ambiguous word is encountered during a sentence and/or this effect may be masked by additional between-sentence integration processes that may occur in ambiguous and unambiguous sentences. However, the nature of its potential effect on LIFG involvement is unclear. Thus, in light of these explanations, the current set of these neuroimaging results appears consistent with the hypothesis that the LIFG is involved in semantic reinterpretation.

The precise nature of the LIFG's involvement in this process, however, is uncertain because there are notable differences between the reinterpretation-related findings reported across studies. The most apparent difference is the locus and extent of the LIFG response. Rodd et al. (2012) report a large cluster of reinterpretation-elevated activation that extends across all three anatomical sub-divisions of the LIFG, with activation centred in posterior pars triangularis. Mason and Just (2007) also report increased activation in a large region of the LIFG when they compared late-disambiguation sentences to unambiguous sentences. However, the peak co-ordinate was much more medial than that reported in Rodd et al. (2012) (x = -28 versus x = -52, respectively) and the cluster appears confined primarily to pars triangularis and pars opercularis, with little anterior spread into pars orbitalis. Yet, the direct comparison between biased and balanced sentences was only localised to a small cluster of activation in an anterior ventral region of the LIFG near the border of pars triangularis and pars orbitalis. In contrast, Zempleni et al. (2007) report activation in a small region of the posterior LIFG only, around the border of pars triangularis and pars opercularis. What can explain these discrepancies? A simple statistical explanation is that these differences are merely a thresholding artefact. Both Zempleni et al. (2007) and Mason and Just (2007) who report less extensive activation than Rodd et al. (2012) do not discuss the spread of activation at lower thresholds than their initial threshold choice. Additionally, Mason and Just (2007) employed a small number of trials (N = 9) per condition and a relatively small sample of subjects (N = 12), which may have reduced sensitivity to detect LIFG responses to reinterpretation demands. Another explanation that is more theoretically interesting is that such differences may reflect inter-subject variability in function-anatomy mappings. An assessment of inter-subject variability of ambiguity-related activation has not yet been investigated. A third possibility is that there may be functional specialisation across the LIFG. For example, activation differences may reflect subtle differences on ambiguity-related processes induced by the employment of different presentation paradigms across the three studies. For example, Rodd et al. (2012) presented sentences in the spoken modality whereas the sentences were presented visually by Zempleni et al. (2007) and Mason and Just (2007) and, furthermore, Zempleni et al. (2007) presented each phrase of their sentence in isolation whereas Mason and Just’s (2007) sentences were presented word-by-word in a cumulative manner. However, it is unclear how such methodological differences would produce this specific pattern of results.

Moreover, it is uncertain how reliable the reported effects of dominance on LIFG activation are, as no direct replications have been published. Although both Zempleni et al (2007) and Mason and Just (2007) examined dominance effects, the former compared subordinate-biased with dominant-biased sentences while the latter compared subordinate-biased with balanced sentences. In addition, both studies employed visually-presented sentences. Thus, whether such patterns replicate for spoken sentences needs to be investigated before being able to generalise these conclusions.

Examining how LIFG responses are modulated by dominance is important for two reasons. First, dominance effects are one of the main predictions of the reinterpretation hypothesis because dominance is a measure of meaning preference and, thus, an index of how likely it is that the contextually inappropriate meaning will initially be preferred and need reinterpretation later on. More specifically, if a region supports semantic reinterpretation then it should respond more to subordinate-biased late-disambiguation sentences than any other ambiguous sentences, since these sentences have a higher probability that the inappropriate (dominant) meaning will initially be selected. Second, up until this point, effects of dominance on neural responses to late-disambiguation sentences have been discussed solely in relation to the reinterpretation hypothesis. However, they are also predicted by an initial meaning selection account but with a different pattern according to current psycholinguistic accounts of ambiguity resolution (Duffy et al., 2001; Twilley & Dixon, 2000). In late-disambiguation sentences, because ambiguous words are encountered prior to contextual information, balanced words place more demands on initial selection processes because the listener has no reason to strongly prefer either meaning compared with biased words for which a single meaning can easily be selected based on frequency information. Eye-movement research provide behavioural evidence of their increased processing demand, finding that readers fixate longer on balanced words than biased words in these kinds of sentences (Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989). Thus, if a region is primarily involved in initial meaning selection then activation should be greater for balanced than biased words in these kinds of sentences. Mason and Just's (2007) dominance pattern clearly show evidence for the reinterpretation hypothesis, as it was the biased sentences that produced greater activation than the balanced sentences. However, no other studies have examined these two conditions. Moreover, Rodd et al.'s (2012) study demonstrated that the LIFG may be involved in both initial meaning selection and reinterpretation processes. Thus, since balanced sentences load more highly on the former and biased sentences load more highly on the latter, then an alternative possibility is that no difference between these two kinds of ambiguous words may be found when an experimental design is used that does not separate out these two processes. Clearly, further examinations of the LIFG response to late-disambiguation sentence are needed, having the potential to elucidate functional contributions of ambiguity-responsive regions.

3.1.3 Theories of the LIFG

Several accounts of LIFG function explain its role in semantic ambiguity resolution.

Conflict Resolution

One theory directly relevant to semantic ambiguity processing is the conflict resolution account (Novick, Kan, Trueswell, & Thompson-Schill, 2009; Novick et al., 2005; Thompson-Schill et al., 1997). According to this theory, the LIFG helps resolve competition between activated representations to support the selection of a single representation. This theory follows the bias competition framework whereby the prefrontal cortex provides top-down signals to bias information processing in other brain regions to support context-appropriate behaviour (Desimone & Duncan, 1995; Miller & Cohen, 2001). Conflict resolution is one of the core processes in psycholinguistic models of ambiguity resolution (e.g., Duffy et al., 2001; Twilley & Dixon, 2000).

Novick et al. (2009) have recently specified that the LIFG supports representational selection in two types of situations: when representational conflict is caused by a prepotent but irrelevant response as well as when multiple representations are available but no dominant response exists. Elegantly, the former seems to map onto semantic reinterpretation demands of ambiguity resolution, where the dominant preference needs to be suppressed in favour of the less frequent meaning, whilst the latter seems to map onto initial selection demands for balanced words which have no strongly dominant meaning. Strong support for the LIFG's role in both types of conflict resolution comes from numerous neuroimaging studies that have employed a wide variety of tasks and stimuli. These include verb generation or picture naming, where stimuli are associated with several possible options (Kan & Thompson-Schill, 2004a; Persson et al., 2004; Thompson-Schill et al., 1997), semantic classification and comparison tasks in which the stimuli evoke strongly associated responses that are irrelevant to the task (Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005; Thompson-Schill et al., 1997), working memory tasks that elicit proactive interference between items in memory such as item recognition tasks (D'Esposito, Postle, Jonides, & Smith, 1999; Jonides & Nee, 2006) and the Stroop task (January, Trueswell, & Thompson-Schill, 2009; Milham, Banich, & Barad, 2003). Lesion studies further corroborate these findings by showing that patients with damage to the LIFG are impaired on similar semantic and non-semantic tasks that have high representational conflict (Hamilton & Martin, 2005; Novick et al., 2009; Robinson, Blair, & Cipolotti, 1998; Robinson, Shallice, & Cipolotti, 2005; Thompson-Schill et al., 2002; Thompson-Schill et al., 1998).

Furthermore, this region is argued to resolve competition across multiple domains of linguistic representation, although Novick et al. (2009) acknowledge the possibility of regional specialisations across domains. Such domain-generality is supported by numerous neuroimaging studies and research on patients that show increased LIFG activation or impaired performance respectively when competition is manipulated within semantic, syntactic or phonetic information, during word and sentence tasks and during both language production and language comprehension (Bedny, McGill, & Thompson-Schill, 2008; Blumstein, Myers, & Rissman, 2005; Fiebach, Vos, & Friederici, 2004; Gennari, MacDonald, Postle, & Seidenberg, 2007; Humphreys & Gennari, 2014; Kan & Thompson-Schill, 2004b; Katzev, Tüscher, Hennig, Weiller, & Kaller, 2013; Novick et al., 2009; Ojanen et al., 2005; Rodd, Longe, et al., 2010; Thompson-Schill et al., 1997). It is also argued to be domain-specific with regard to the type of conflict, specifically attributed to conceptual rather than response-based conflict and to verbal as oppose to non-verbal information. These dissociations are supported by the finding that tasks which manipulate response competition, such as Go/No-Go and Stop Signal tasks, produce consistent activation in right, rather than left, hemisphere frontal regions (see Nee, Wager, & Jonides, 2007 for a meta-analysis) and patients with damage to the LIFG show normal performance on nonverbal, but not language, tasks that have high conflict demands (Hamilton & Martin, 2005; Robinson et al., 2005).

Moreover, Novick et al. (2009) argue that the posterior region of the LIFG (BA44/45) is the critical site for representational conflict resolution. An extensive fMRI study conducted by Badre and Wagner (2005) found evidence supporting such a posterior locus for this function in the LIFG that dissociated from controlled semantic retrieval processes attributed to the anterior LIFG (discussed in more detail later). This suggests that different sub-regions of the LIFG may be serving different functional roles in ambiguity resolution.

Another distinguishing claim of this theory with regard to language processing is that the posterior LIFG is engaged only when there is representational conflict. This is supported by a body of research showing that the LIFG is not necessary for language comprehension in the absence of such conflict. For example, patients with “Broca’s aphasia” or damage to the LIFG itself have relatively preserved comprehension of single words and simple sentences (Caplan, Hildebrandt, & Makris, 1996; Caramazza & Zurif, 1976; Grodzinsky, Piñango, Zurif, & Drai, 1999; Schwartz et al., 1980; Yee, Blumstein, & Sedivy, 2008) and can produce and comprehend sentences that have low conflict demands such as when there is a dominant response (Novick et al., 2009; Robinson et al., 2005). Lesion-deficit analyses have also failed to find significant correlation between damage to the LIFG and single word or sentence comprehension (Bates et al., 2003; Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004). In addition, neuroimaging studies do not always report significantly greater activation in the LIFG for simple or unambiguous sentences over baseline (e.g., Crinion et al., 2003; Rodd, Johnsrude, et al., 2012; Scott, Blank, Rosen, & Wise, 2000; Spitsyna, Warren, Scott, Turkheimer, & Wise, 2006).

Controlled Semantic Retrieval

Badre and colleagues specifically attribute the anterior LIFG (pars orbitalis) to controlled semantic retrieval (Badre et al., 2005; Badre & Wagner, 2007). According to this theory, this region supports the controlled retrieval of semantic information from long-term memory. This process is required when stimulus-driven cues are insufficient to activate information relevant to one’s goal or task, such as in the case of a semantic decision where two stimuli are weakly rather than strongly associated. This process, therefore, seems particularly important to resolve ambiguous sentences that refer to the subordinate meaning of an ambiguous word, as these meanings are more weakly associated to the ambiguous words than their dominant meaning. Badre and Wagner (2005) specifically localised controlled semantic retrieval to the anterior LIFG, based on their extensive fMRI study that dissociated this function from selection processes by employing various tasks that loaded more heavily on one of these processes than the other. A recent TMS study provides direct support for a causal role of the LIFG in controlled semantic retrieval by showing that stimulating the LIFG impaired performance on semantic decision tasks for conditions with weak, but not strong, semantic associates (Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011).

Semantic Unification

Hagoort and colleagues propose an alternative language-general account of the LIFG function that contrasts with the stimulus-specific views taken in the conflict resolution and controlled semantic retrieval accounts (Hagoort, 2005, 2013; Hagoort, Baggio, & Willems, 2009). According to this theory, the LIFG constitutes a unification space that combines basic units of information to form larger complex representations of a sentence or discourse. In contrast to the conflict resolution account, this theory argues that combinatorial demands drive activation in this region rather than conflict or selection load per se. Selection is merely one aspect of unification. This claim is supported by the findings of several neuroimaging studies which show greater activation in the LIFG for sentences without conflicting information, such as semantically congruent or semantically unambiguous sentences, compared to rest or a low-level baseline (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Rodd et al., 2005; Willems, Özyürek, & Hagoort, 2007) as well as for sentences compared with random word lists which are assumed not to require unification into a higher-level meaning (Snijders et al., 2009). As discussed previously, various other neuroimaging studies have found null-results for such non-conflict sentences, which have been taken as evidence for conflict-specific accounts of the LIFG. However, null-results may merely be due to increased activation in baseline conditions that may mask significant activation in language conditions (e.g.,

J. Binder et al., 1999). In addition, this theory can account for some empirical findings that are not straightforward to explain with the conflict resolution theory. For example, the LIFG shows increased activation for sentences that contain linguistic violations, such as semantically or syntactically anomalous words (Baumgaertner, Weiller, & Büchel, 2002; Hagoort et al., 2004; Kuperberg, Sitnikova, & Lakshmanan, 2008; Zhu et al., 2012) compared with normal sentences. Although there is conflict between representations in anomalous sentences, they do not require selection processes to resolve the conflict (i.e., the LIFG function of the conflict resolution account). On the other hand, there is clearly a greater load on unification processes.

The unification account also explicitly proposes functional specialisation across the LIFG in terms of the type of linguistic information that is processed. Based on a literature review by Bookheimer (2002), Hagoort (2005) argues that different types of information are processed between large overlapping regions. Specifically, semantic information is preferentially processed by pars orbitalis and pars triangularis, syntactic information by pars triangularis and par opercularis and phonological processes by pars opercularis and premotor cortex. Such a division has recently been supported by a resting state study showing different patterns of functional connectivity between the three sub-regions, suggesting that they form parts of different information processing streams (Xiang, Fonteijn, Norris, & Hagoort, 2010). This fractionation is not entirely compatible with that argued by the conflict resolution account, which claims that the posterior, not the anterior, LIFG is the critical site for resolution between semantic representations. Thus, although Hagoort and colleagues explicitly incorporate semantic selection operations within the set of unification functions supported by this anterior region (Hagoort, 2005; Hagoort et al., 2009), it is unclear whether this is appropriate given that high-conflict semantic tasks frequently show posterior activation. Nevertheless, numerous other researchers support the notion of functional specialisation across the LIFG, although the exact type of operations performed in these sub-divisions differ across accounts (Badre & Wagner, 2007; Friederici, 2012; Gough, Nobre, & Devlin, 2005; Poldrack et al., 1999; Vigneau et al., 2006). For example, the controlled semantic retrieval account described above is a more specific semantic-based theory that is associated with the anterior LIFG (Badre & Wagner, 2007), while various phonological processes have been attributed to the posterior LIFG, described further below.

Phonological Accounts of the LIFG

A variety of phonological processes have been associated with the posterior LIFG, namely pars opercularis. Aside from phonological unification (Hagoort, 2005, 2013), others include phonological working memory processes specifically associated with maintaining phonological representations (Nixon, Lazarova, Hodinott-Hill, Gough, & Passingham, 2004), articulatory operations (Hickok & Poeppel, 2007) and phonological sub-vocal rehearsal (Rogalsky, Matchin, & Hickok, 2008). These theories have received support from a variety of neuroimaging, TMS and lesion studies, employing paradigms such as letter/word rhyming judgements (Baldo & Dronkers, 2006; Gough et al., 2005), delayed phonological matching of words/pseudowords (Nixon et al., 2004; Strand, Forssberg, Klingberg, & Norrelgen, 2008), concurrent comprehension and articulation (Rogalsky et al., 2008) and phoneme detection tasks (Zatorre, Evans, Meyer, & Gjedde, 1992).

Most cognitive theories of semantic ambiguity resolution do not highlight phonological-based mechanisms as mediating this process. However, there is reason to hypothesise that semantic reinterpretation, in particular, may involve these kinds of operations. For example, Rodd et al. (2012) propose that phonological working memory may be specifically necessary to reprocess sentences that have been misunderstood so that previous parts of the sentence can be re-evaluated into a new coherent interpretation. Rogalsky et al. (2008) suggest a similar, but more explicit, mechanism with regard to processing syntactically complex sentences, proposing that, when listeners hear complex sentences (i.e., object-relative sentences), they may use articulatory rehearsal to repeat back what they have heard to check that their analysis is correct. It is also possible that initial meaning selection may also engage phonological re-activation processes to support the appropriate selection of a single meaning. However, this is likely to be less demanding than that which occurs for semantic reinterpretation, as less of the sentence would have been heard at the time that the ambiguous word is encountered than when reinterpretation is needed (i.e., at the end of sentence).

In summary, several theories of LIFG function predict involvement of this region in resolving the meaning of semantically ambiguous words during sentence comprehension. Although the specific functional and anatomical details differ between the accounts, it is clear that all of these theories predict that, some, if not all, sub-divisions of this region, should be engaged when sentences require semantic reinterpretation. Thus, all theories predict that the LIFG should show greater activation for late-disambiguation sentences compared with low-ambiguity control sentences. Some theories also predict its involvement in initial meaning selection (i.e., unification and conflict resolution accounts). Since the theories suggest different involvements in these two components of ambiguity resolution (initial meaning selection and reinterpretation), they make different predictions about how these regions' responses may be modulated by meaning dominance. Theories that only suggest a role in semantic reinterpretation predict that biased sentences should show greater activity than balanced sentences because it is the biased sentences that have the greatest likelihood of needing reinterpretation. In contrast, theories that suggest roles in both initial selection and semantic reinterpretation predict that balanced sentences may produce no greater activation than biased sentences because balanced meanings are hardest to select between, and, thus, may equate to the neural demands produced for reinterpretation in the biased sentences. No theories currently suggest an exclusive role for the LIFG in initial selection processes but not in reinterpretation, and thus, no theories predict that balanced sentences should produce greater activity than biased sentences. Current evidence cannot distinguish between these accounts.

Left Posterior Temporal Cortex and Semantic Ambiguity Resolution

The left posterior inferior temporal cortex is a second region that is often activated alongside the LIFG in response to semantically ambiguous sentences (Davis et al., 2007; Rodd et al., 2005; Rodd, Johnsrude, et al., 2012; Zempleni et al., 2007). Its ambiguity response appears more discriminate than that of the LIFG, showing robust activation for only certain types of ambiguous sentences (Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007), which suggests that this region may play a functionally different role in ambiguity resolution compared to frontal regions. For example, Rodd et al. (2012) argue that, unlike frontal regions, the inferior temporal lobe has “a relatively pure contribution to reinterpretation processes”. This conclusion was based on the results of their time course study, which showed that posterior inferior temporal regions responded to disambiguating information when it was encountered after, but not before, an ambiguous word and did not show greater activation to the ambiguous word itself. Like the frontal region, its involvement in semantic reinterpretation is corroborated by the finding that meaning dominance has been shown to affect responses in the temporal cortex. Specifically, Zempleni et al. (2007) showed greater activity in the posterior middle/inferior temporal gyrus for subordinate than dominant-biased sentences. The specificity of this region’s role in reinterpretation is consistent with other reports of posterior inferior activation for late-disambiguation sentences (Bekinschtein et al., 2011) as well as with the lack of ambiguity-elevated activation reported for sentences in which reinterpretation is unlikely, such as when the disambiguating information is presented immediately after an ambiguous word (Rodd, Longe, et al., 2010) or for dominant-biased sentences (Zempleni et al., 2007). Not all studies, however, have found such supporting evidence. For example, Mason and Just (2007) do not report increased activation anywhere in the temporal lobe for their set of late-disambiguation sentences compared with unambiguous sentences, even when biased sentences, which have the highest likelihood of reinterpretation, were analysed separately to balanced sentences. Yet, this may result from a lack of power as only nine sentences were used per condition and 12 subjects were tested. In addition, the scanning protocol may have been sub-optimal for detecting activation in the inferior temporal region since the onset of the ambiguous word was synchronised to the acquisition of the most superior axial slice.

As with frontal activation, there is variability in the location and extent of ambiguity-elevated activation reported in the posterior temporal cortex. Activation is not confined to a specific anatomical region but often encompasses several functionally distinct temporal regions. All studies report responses in the posterior inferior temporal gyrus (Bekinschtein et al., 2011; Davis et al., 2007; Rodd et al., 2005; Rodd, Johnsrude, et al., 2012; Zempleni et al., 2007), with some studies also reporting activation in the posterior middle temporal gyrus (Rodd et al., 2005; Rodd, Johnsrude, et al., 2012; Zempleni et al., 2007) and/or ventrally within the fusiform gyrus (Bekinschtein et al., 2011; Rodd et al., 2005; Rodd, Johnsrude, et al., 2012). One study even found activation that extended more superiorly into the posterior superior temporal sulcus (STS) (Rodd et al., 2005). Again, these discrepancies may reflect effects of statistical thresholds, inter-subject functional variability or differences in ambiguous stimuli or experimental protocols.

3.1.4 Theories of the Posterior Temporal Cortex

The role of the posterior temporal cortex in semantic ambiguity resolution is controversial due to the fact that widely different accounts have been proposed for its function in language and cognition in general. Unlike LIFG theories that converge on cognitive control functions, the posterior temporal cortex has been attributed to a more heterogeneous set of processes. Three distinct accounts relate to sound-meaning mapping (Hickok & Poeppel, 2004, 2007), semantic control (Jefferies, 2013) or visual-based semantic information (Cohen, Jobert, Le Bihan, & Dehaene, 2004; Martin, 2007; Price & Devlin, 2003, 2011).

Sound-meaning Interface

Hickok and Poeppel propose that the middle and inferior portions of the posterior temporal lobe constitute a lexical interface that supports the mapping between phonological and semantic representations (Hickok & Poeppel, 2004, 2007). A recent version of this theory (Hickok & Poeppel, 2007) emphasises the left posterior MTG (pMTG) and neighbouring inferior temporal sulcus (pITS) as the critical sites in this process. Less focus is given to the inferior temporal cortex than in earlier versions (Hickok & Poeppel, 2004). This theory was initially based on the existence of an aphasic syndrome known as transcortical sensory aphasia that presents symptoms suggestive of a deficit in sound-meaning mapping. Such patients have impaired auditory comprehension with relatively intact syntactic and phonological processing and damage is typically observed in the posterior inferior temporal lobe (A. R. Damasio, 1991; Kertesz, Sheppard, & MacKenzie, 1982). This hypothesis has been supported by several other lines of research. Two large-scale lesion studies have found significant associations between pMTG damage and auditory comprehension deficits (Bates et al., 2003; Dronkers et al., 2004). Direct cortical stimulation of posterior middle and inferior cortex produces speech and auditory comprehension problems (Boatman et al., 2000; Lüders et al., 1991). In addition, patients with semantic dementia, although typically associated with anterior temporal damage, show hypo-perfusion of left posterior temporal areas which also show reduced activation during semantic judgement tasks (Garrard & Hodges, 2000; Mummery et al., 1999). Furthermore, numerous fMRI studies show increased activation in pMTG in response to routine lexical-semantic processes such as during auditory sentence comprehension (Davis et al., 2007; Rodd et al., 2005; Rogalsky &

Hickok, 2009; Saur et al., 2008) and spoken word recognition (Davis & Gaskell, 2009). Additionally, word and picture naming, for which sound-meaning mapping seems especially important, also produce activation in posterior middle and inferior temporal regions (Indefrey & Levelt, 2004; Price, 2012; Price, Wise, Warburton, et al., 1996). It is currently uncertain why ambiguous words would engage sound-meaning mapping processes imputed to this region. Rodd et al. (2012) suggested that these kinds of representations may be reactivated in the reinterpretation process when the meaning of an ambiguous sentence needs to be recomputed. However, there is currently no behavioural research that has tested such a theory.

There are several aspects of the discussed evidence that suggest that this theory of the posterior temporal region's involvement in language comprehension, and thus ambiguity resolution, may be inadequate. First, several observations query the localisation of a sound-meaning mapping function to the posterior middle and inferior temporal structures. Although activation in auditory sentence comprehension and spoken word recognition studies is often found in the pMTG, the corresponding cluster often forms part of a superior temporal cluster with the posterior STG and/or extends along the MTG, rather than extending into inferior temporal regions in the way that is typically seen for studies of ambiguity resolution (Bekinschtein et al., 2011; Davis et al., 2007; Rodd et al., 2005; Zempleni et al., 2007). Some other studies of sentence comprehension do not even report any activation in the pMTG, instead, showing activation restricted to the posterior STG/STS (Adank & Devlin, 2010; Humphries, Willard, Buchsbaum, & Hickok, 2001; Obleser & Kotz, 2010) and/or mid-anterior portions of the MTG (Humphries et al., 2001). Thus, these findings demonstrate that sound-to-meaning mapping does not always engage the pMTG/ITS. Hickok and Poeppel's (2007) current account does not attempt to explain these discrepancies. It is possible to reconcile these findings with their account if sound-to-meaning mapping is not viewed as a unitary process. Under this view, the pMTG/ITS may serve functionally different operations that facilitate the mapping between phonological and sound information compared to those served by the more superior temporal and anterior regions. Since the former regions are more consistently found during ambiguous sentence comprehension than during unambiguous sentence comprehension (compared to a low-level baseline) it may be serving more high-level mapping processes than those that occur during more normal sentence comprehension (Rodd et al., 2005). More research is needed that directly examines the extent to which the same temporal regions are engaged by different demands on sound-meaning operations.

Recent lesions studies, however, have reported more problematic findings for Hickok and Poeppel's (2007) account of this region. Specifically, they have called into question one of the fundamental aspects of this theory, that this region processes verbal information. These studies have found that patients with damage to posterior temporal areas show deficits on both verbal and nonverbal semantic tasks (Corbett, Jefferies, & Lambon Ralph, 2009; Corbett, Jefferies, & Lambon Ralph, 2011; Robson, Sage, & Lambon Ralph, 2012). Nonverbal deficits cannot be explained by a lexical-focused account. It is, as yet, unclear exactly how to accommodate these findings within those of Hickok and Poeppel's (2007) account. For example, the posterior temporal region may perform multiple functions or it may serve a more general function that is involved in tasks of sound-meaning mapping and nonverbal semantic tasks.

Semantic Control

Jefferies and colleagues argue for an alternative theory of the pMTG that explicitly accounts for the nonverbal deficits of patients with lesions to this region (Jefferies, 2013; Noonan, Jefferies, Visser, & Lambon Ralph, 2013; Whitney, Jefferies, & Kircher, 2011). According to Jefferies (2013), pMTG supports processes associated with semantic control that act to enhance task-relevant activation within the semantic system. This theory is supported by the pattern of deficits shown by patients with semantic aphasia who have damage to posterior temporal regions (as well as inferior parietal and inferior frontal regions). Such patients are impaired on a wide variety of semantic tasks across multiple input and output modalities, including spoken and written words, pictures, naming, alternative forced choice, nonverbal tasks (Corbett, Jefferies, Ehsan, & Lambon Ralph, 2009; Corbett, Jefferies, & Lambon Ralph, 2009; Corbett et al., 2011; Jefferies, Baker, Doran, & Lambon Ralph, 2007; Jefferies & Lambon Ralph, 2006; Jefferies, Patterson, & Lambon Ralph, 2008; Noonan, Jefferies, Corbett, & Lambon Ralph, 2010). However, their performance is inconsistent across tasks, being particularly affected by semantic demands of the task. For example, these patients benefit from response cues during picture naming (Jefferies et al., 2008), and are also affected by semantic distracters that are irrelevant to the task (Noonan et al., 2010). Jeffries (2013) specifically argue that these effects derive from a deficit in semantic control processes, since semantic demands are lowered by response cues and increased by semantic distracters.

The localisation of this function to pMTG is corroborated by many neuroimaging studies, showing increased activation in this region for semantic control tasks that were aforementioned with regard to the LIFG (e.g., Badre et al., 2005; Gold & Buckner, 2002; Noppeney, Phillips, & Price, 2004; Thompson-Schill et al., 1997; Whitney, Jefferies, et al., 2011). More recent causal evidence has been demonstrated in a recent TMS study, where stimulating this region affected performance on tasks that had high, but not low, semantic control demands (Whitney, Kirk, et al., 2011) as well as by a recent patient study on Wernicke's aphasia, finding that semantic processing deficits were associated with damage to this region whereas phonological deficits were associated with pSTG damage (Robson et al., 2012). Such a theory is also consistent with the conclusions of other researchers that argue for semantic integration as one key function of this region, based on the finding that it is more active when demands on integration are high such as for sentences with semantic anomalies and ambiguities (Baumgaertner et al., 2002; Zempleni et al., 2007). In addition, the semantic control theory can also explain why this region is not always activated in language studies, as it depends on the semantic demands of the task.

The role of posterior temporal cortex in semantic control is a relatively recent theoretical perspective but seems to fit with the coupling of response with the LIFG and provides a straightforward explanation for its involvement in ambiguity resolution.

Visual-based Semantic Information

Both of the above theories strongly diverge from the traditional view of the posterior temporal cortex as a high-level visual association area. The posterior temporal cortex is located anterior to amodal visual association areas of the occipital lobe and the lateral and ventral occipitotemporal cortex, whose strong responses to visual input have been firmly established by monkey and human research (Grill-Spector & Malach, 2004). Visual-based accounts of these regions lead to a radically different perspective of the roles they may serve in ambiguity resolution and language comprehension more generally. According to one view (Martin, 2007), these posterior regions represent conceptual information about the visual features of objects, with different types of information being stored in different regions. Martin (2007) views semantic representations as grounded in similar networks associated with perceptual processing of objects, such that understanding the meaning of word that corresponds to a concrete object (e.g., fan) necessarily involves activating its visual features. Thus, according to this theory, these regions are more activated during ambiguous sentences because such words place greater demands on semantic representations.

Accordingly, in strong contrast with the two previous theories, the pMTG is associated with object knowledge and more specifically with motion-related attributes of objects. This semantic role of pMTG has received a large amount of support from neuroimaging research, with studies showing greater activation in this region for motion-related stimuli, such as point-light displays (Beauchamp, Lee, Haxby, & Martin, 2003) as well as for words or pictures of tools than other stimuli such as animals (e.g., Chao, Weisberg, & Martin, 2002; H. Damasio, Grabowski, Tranel, & Hichwa, 1996; Devlin et al., 2002; Devlin, Rushworth, & Matthews, 2005; Moore & Price, 1999), for verbs more than nouns or prepositions (e.g., H. Damasio et al., 2001; Kable, Lease-Spellmeyer, & Chatterjee, 2002; Perani et al., 1999), and its activation has been found to predict recall of names of manipulable objects

(Polyn, Natu, Cohen, & Norman, 2005). Various lesion studies have also provided direct support for its causal role in action representations (Tranel, Damasio, & Damasio, 1997; Tranel, Kemmerer, Adolphs, Damasio, & Damasio, 2003). Its involvement in visual motion information is further strongly supported with the fact that it is located just anterior to MT which is the primary visual motion processing region. Gennari et al. (2007) provide supporting evidence that this region's involvement during ambiguity resolution may be explained by its role in motion-related semantics. In this study, participants read noun or verb phrases containing an ambiguous word (e.g., "the bowl" versus "to bowl"). fMRI results showed that pMTG had greater activity for ambiguous than unambiguous phrases, supporting its involvement in ambiguity processing but, critically, this region was also affected by the syntactic context, with verbs producing greater activity than nouns. However, this account cannot entirely explain the finding of ambiguity-responses in this region when ambiguity is manipulated between two noun definitions rather than between a verb and noun (Rodd et al., 2005; Rodd, Johnsrude, et al., 2012). The more inferior posterior temporal regions, commonly referred to as ventral occipitotemporal cortex (vOT), have also been attributed to object representation but to the visual form of an object rather than motion-related attributes. This conclusion is based on evidence that different areas show specificity for certain types of concrete objects (Martin & Chao, 2001), is affected by semantic priming of object words (Wheatley, Weisberg, Beauchamp, & Martin, 2005) and are active when participants view these objects as well as when they imagine them (Ishai, Ungerleider, & Haxby, 2000; O'Craven & Kanwisher, 2000).

Other lines of research, however, demonstrate evidence for a processing gradient in vOT, which is not accounted for in the general visual form theory. Such a gradient has various implications for the involvement of this region in ambiguity resolution and language comprehension in general. Based on a wealth of neuroimaging studies as well as more recent functional and structural connectivity work, Price and Devlin (2011) argue for a posterior-anterior gradient of information abstractness across the occipitotemporal cortex, with a three-way division of labour: the posterior portion ($y < -60$ in MNI space) is primarily related to visual input, the middle portion ($-60 > y > -50$) integrates visual and non-visual (semantic, phonological) information while the anterior portion ($y > -50$) is associated with multimodal semantic processes. Posterior vOT is critically not associated with specific object features per se but is driven by high-level visual processing in general. Support for this comes from the finding that this region shows greatest activation for an array of visual input, regardless of the meaningfulness of the stimulus (Price, Moore, Humphreys, Frackowiak, & Friston, 1996; Price, Noppeney, Phillips, & Devlin, 2003; van Turennout, Ellmore, & Martin, 2000), is rarely found to respond to non-visual stimuli such as auditory words or tactile stimulation (J. Binder, Desai, Graves, & Conant, 2009; Kassuba et al., 2011; Price, Winterburn, Giraud, Moore, & Noppeney, 2003) and lateralization is related to visual feature information (Seghier & Price, 2011). In line with this, only one fMRI study on semantic ambiguity resolution has reported activation so posteriorly in this region (Bekinschtein et al., 2011: peak at $y = -62$). More often, ambiguity-elevated activation is reported within the anterior-middle vOT (Rodd et al., 2005: $y = -40$; Bekinschtein et al., 2011: $y = -48$; Rodd et al., 2012: $y = -38$, $y = -50$). The claim that the anterior vOT supports multimodal semantic processing is supported by various neuroimaging studies showing increased activation for visual, auditory and tactile stimuli and for semantic processing more than phonological processing in this region (Davis & Gaskell, 2009; Kassuba et al., 2011; Price, Winterburn, et al., 2003; Sharp et al., 2010). Its role in heteromodal semantic processing has been supported further by a recent extensive meta-analysis that showed overlap in anterior fusiform across neuroimaging studies of semantics (J. Binder et al., 2009). However, the exact role this region may serve in semantic processing is as yet uncertain.

The role of middle vOT in ambiguity resolution is less straightforward. According to Price and Devlin (2011) this region acts as an interface between visual and language areas representing non-visual information such as semantics and phonology. This theory is supported by the finding that this region shows greater response to visual than non-visual stimuli (Price, Winterburn, et al., 2003) but, unlike posterior vOT, is also modulated by object meaningfulness (Castelli, Frith, Happé, & Frith, 2002; van Turennout et al., 2000) and language demands on the task, such as when semantic or phonological information interferes with or facilitates lexical decisions (Devlin, Jamison, Gonnerman, & Matthews, 2006; Twomey, Kawabata Duncan, Price, & Devlin, 2011). The influence of these factors on processing in this region is argued to reflect a top-down influence of language areas on visual information processing.

Other researchers postulate a more specific role of this region in language, where it specifically represents orthographic features of words (Cohen et al., 2004; Dehaene & Cohen, 2011; Kronbichler et al., 2004). This is based on the consistent findings that this region shows greater activation for visual words and pseudowords than a number of other visual stimuli (e.g., false fonts, checkerboards) (Cohen et al., 2002) and is affected by word frequency (Kronbichler et al., 2004). Neither of these theories simply explains nor predicts its involvement in ambiguity resolution as they both emphasise the visual nature of information processing in this region. However, like the explanation based on Martin's (2007) object account, ambiguity-related activation may reflect top-down activation of visual information related to a word's meaning, which is consistent with the regions postulated interaction with the language system (Price & Devlin, 2011) or top-down activation of the orthographic representation of the word (Dehaene & Cohen, 2011).

In summary, neural investigations of semantically ambiguous sentences consistently report activation within the posterior middle and inferior temporal cortex. Such activation has been found over a set of temporal regions that have been proposed to have different functions. My review of these regions shows that certain regions, such as pMTG and anterior vOT, are strongly predicted to be involved during ambiguity resolution. Further work is evidently needed to elucidate the consistent regions within the network underlying this process.

3.1.5 Present Study

The current literature raises several questions regarding the involvement of the LIFG and posterior temporal cortex in semantic ambiguity resolution. What are the specific anatomical sub-fields engaged by semantic ambiguity? What roles do they play? How consistent is this ambiguity-network across individuals?

These questions were investigated using fMRI. Neural responses to a large set of late-disambiguation sentences were compared with those to well-matched unambiguous sentences. Based on previous research, it was predicted that ambiguity-elevated responses would be broadly found in the LIFG and the left posterior temporal cortex (Bekinschtein et al., 2011; Davis et al., 2007; Mason & Just, 2007; Rodd et al., 2005; Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007).Various theories of LIFG and posterior temporal cortex function predict their involvement in ambiguity resolution. However, they differ in terms of the specific anatomical region that is described. The roles of ambiguity-responsive regions were assessed further by examining 1) their responses to unambiguous sentences (compared to a low-level baseline) and 2) their modulation by dominance. The “unambiguous contrast” provides an assessment of the nature of these roles, specifically whether they reflect operations that are routinely involved in sentence comprehension or are more specific to ambiguous stimuli. Only theories that advocate a general language mechanism predict greater activation for unambiguous sentences than baseline such as the semantic unification account of the LIFG (Hagoort, 2005, 2013), and the lexical-semantic processing (Hickok & Poeppel, 2007) and object representation accounts of the posterior temporal cortex (Martin, 2007). On the other hand, explicit proponents of a non-language-general mechanism are the LIFG conflict resolution account (Novick et al., 2009; Novick et al., 2005; Thompson-Schill et al., 1997) and the pMTG semantic control account (Jefferies, 2013). These do not predict responses to unambiguous sentences (compared to a low-level baseline).

The dominance analysis examines the contribution of these regions to the two key functions of ambiguity resolution, initial meaning selection and semantic reinterpretation, by comparing responses to biased and balanced sentences. Critically, like Mason and Just (2007), the biased sentences were always disambiguated toward their subordinate meaning. If these regions are primarily involved in semantic reinterpretation, then responses should be greater for biased than balanced sentences because they have a stronger frequency bias that makes the inappropriate (dominant) meaning likely to be initially selected on most trials and thus need reinterpretation. In contrast, any regions that are primarily involved in initial semantic selection should show greater activation for balanced than biased sentences, because this process is more difficult since listeners have less strong preferences for the alternative meanings of these words. If regions are involved in both processes (relatively equally), then they should show equivalently strong activation to both balanced and biased sentences, since they both load on (at least) one of these processes. Two types of biased sentences were also compared: strongly-subordinate and weakly-subordinate words. This comparison allows us to examine whether responses are merely related to the likelihood of reinterpretation, where the dominance pattern would be: “strongly biased” > “weakly biased” > “balanced”, or whether a less linear relationship exists between reinterpretation and ambiguity-responses. For example, a region may be especially engaged when very infrequent meanings need to be integrated, which would produce a pattern of: “strongly biased” > “weakly biased” = balanced”. This latter comparison has not yet been investigated.

Furthermore, inter-subject variability of these ambiguity responses were assessed by examining whether the regions that showed reliable activation at the group-level were activated in all subjects.

3.2 Method

Participants

20 native monolingual British English speakers, aged 18-35 (M = 23.8) participated in the study. All were right-handed, had normal or corrected-to-normal vision and had no known hearing or language impairment. Participants were recruited via the University College London (UCL) subject pool and were paid for their participation.

Stimuli

92 late-disambiguation sentences and 92 unambiguous sentences were employed. 90 of them were the ones used in the preceding behavioural experiments (Chapter 2). The other four sentences (two ambiguous and two unambiguous) were constructed in the same way but had to be excluded in the behavioural experiments due to constraints on having multiple versions that required sentences to be divisible by three.

The 92 ambiguous words had four types of meaning dominance (see method section in Chapter 2). These categorizations were based on categories reported in the ambiguity literature (Duffy et al., 1988; Rayner & Duffy, 1986; Sereno, 1995; Vuong & Martin, 2011). (1) 32 words were strong subordinate-biased, where the meaning used in the experimental sentences is very infrequent, on average, preferred by only 6% of listeners (dominance range: 0-0.14); (2) 27 words were weak subordinate-biased, where the sentence meaning is fairly infrequent, on average preferred by 21% of listeners (dominance range: 0.16-0.30); (3) 27 words were balanced, where the sentence meaning is one of two (or more) relatively equally frequent meanings, on average, preferred by 39% of listeners (dominance range: 0.31-0.54; (4) six sentences had high dominance scores, where the sentence meaning is, on average preferred by 77% of listeners (dominance range = 0.65-0.84). Only the first three dominance conditions were used to assess the effect of meaning dominance on neural responses, since the fourth, “high-dominance”, condition had too few sentences. One-way independent-measures ANOVAs showed that the strong-biased, weak-biased and balanced conditions did not differ significantly on various sentence-level properties (duration in seconds: F(2,83) = 1.16, p = .317, number of syllables: F(2,83) = 1.60, p = .208, number of words: F(2,83) = .264, p = .768, position of the ambiguous word: F(2,83) = 1.49, p = .232, position of the disambiguating word: F(2,83) = 1.22, p = .301, naturalness rating: F(2,83) = 1.10, p = .337) nor in terms of the ambiguous words’ lexical properties (overall frequency: F(2,83) = 2.45, p = .093, log frequency: F(2,83) = 2.35, p = .102, number of letters: F(2,83) = 1.92, p = .153, number of meanings: F(2,83) = 1.06, p = .351 or number of senses: F(2,83) = .073, p = .930). See Table 3-1 and Table 3-2 for descriptive statistics of these properties of the three dominance conditions.

Table 3-1: Descriptive statistics (mean(SD)) for properties of the ambiguous and unambiguous target words

Target Word	Frequency per million	Log frequency	No. letters	No. meanings	No. senses
Ambiguous (all)	61.7 (78.5)	3.61 (1.01)	4.72 (1.16)	1.92 (0.90)	10.1 (5.60)
(N=32) Strong Subordinate	70.1[a] (81.0)	3.66 (1.16)	4.50 (1.02)	1.81 (0.64)	9.91 (4.79)
(N = 27) Weak Subordinate	78.5[a] (106)	3.82 (1.01)	5.04 (1.43)	1.89 (0.97)	10.4 (7.03)
(N = 27) Balanced	33.7 (31.0)	3.24 (0.81)	4.56 (0.89)	2.15 (1.10)	10.0 (5.06)
Unambiguous	56.4 (62.6)	3.63 (0.93)	4.76 (1.08)	1.09 (0.32)	4.90 (3.09)

[a] Means are inflated by an outlier value such that the median value for the strong-subordinate and weakly-subordinate condition are considerably lower (38.5 and 31.0 respectively).

The 92 ambiguous words had a significantly higher mean number of meanings (t(91) = 8.14, p < 0.001) and senses (t(91) = 8.31, *p* < 0.001) than the unambiguous targets, as indexed in the Online Wordsmyth English Dictionary-Thesaurus (Parks, Ray, & Bland, 1998). There was no significant difference between the ambiguous and unambiguous targets' word frequency in the CELEX lexical database (Baayen et al., 1995) (t(91) = 1.52, p = .13), log-transformed word frequency (t(91) = .35, p = .73) or number of letters (t(91) = .41, p = .68). See Table 3-1 for descriptive statistics of these lexical properties. Additionally, the unambiguous and ambiguous sentences still did not differ statistically in terms of physical duration (t(91) = .02, p = .99), number of syllables (t(91) = 1.49, p = .14) or number of words of the whole sentence. Both sets of sentences had relatively high mean naturalness ratings but the ambiguous sentences were rated as significantly less natural than the unambiguous sentences (*t*(91) = 3.98, *p* < .001). See Table 3-2 for descriptive statistics of these sentence properties.

Table 3-2: Descriptive statistics (mean(SD)) for properties of the ambiguous and unambiguous sentences

Sentence	Length (seconds)	No. syllables	No. words	Naturalness rating
Ambiguous (all)	2.97 (0.29)	16.5 (1.87)	12.5 (1.23)	5.46 (0.62)
(N=32) Strong Subordinate	2.96 (0.25)	16.3 (1.82)	12.6 (1.34)	5.37 (0.68)
(N = 27) Weak Subordinate	3.03 (0.30)	16.9 (1.92)	12.4 (1.18)	5.60 (0.61)
(N = 27) Balanced	2.91 (0.34)	16.1 (1.69)	12.6 (1.15)	5.40 (0.56)
Unambiguous	2.97 (0.31)	16.4 (1.91)	12.5 (1.23)	5.80 (0.61)

Additionally, 46 filler sentences were employed with the same structure as the experimental sentences. 14 were used in an initial practice block, 24 were catch sentences and the remaining 8 were unambiguous sentences that constituted dummy trials at the beginning of the fMRI runs. Catch sentences were to be presented with a visually presented probe word which participants would need to decide was related or unrelated to the sentence. The aim of the catch-trials was to ensure attention was being paid to the sentences. Thus, for each catch sentence, a probe word was selected that was either clearly semantically related (50%) or clearly unrelated (50%) to the sentence’s meaning. No probes were related to the contextually inappropriate meanings of the ambiguous words. All sentences were recorded by the same female speaker.

Furthermore, 32 experimental sentences were randomly selected and converted to signal-correlated noise (SCN) using Praat software (http://www.praat.org). These SCN stimuli constituted the low-level auditory baseline condition. They are created by replacing all the spectral detail with noise, which makes them completely unintelligible whilst retaining the original spectral profile and amplitude to maintain low-level acoustic properties. An additional two sentences were selected and converted to SCN for the practice block.

The auditory stimuli were delivered over Sensimetrics insert earphones (http://www.sens.com/s14/) in the scanner. EQ filtering Software (Sensimetrics, Malden, MA, USA) was used to filter all sound files to ensure accurate frequency reproduction.

Finally, the stimuli were divided into four lists constituting the four runs of the experiment. The stimuli were pseudo-randomised so that each list had an equal number of each stimulus type and no ambiguous sentence was placed in the same list as its matched unambiguous sentence in order to avoid potential syntactic priming effects.

Design and Procedure

An event-related, within-subject design was employed in which participants were randomly presented with all types of sentence trials (ambiguous, unambiguous, SCN and catch sentences) as well as silent (rest) trials lasting, on average, the same physical duration as the sentence trials. Rest trials were included as another baseline condition. The experiment was divided into four sessions, each with 70 trials: 23 ambiguous, 23 unambiguous, 8 SCN and 8 rest trials as well as two dummy trials to allow for T1 equilibrium before the test trials began. Each session lasted, on average, 8.47 minutes. The order of the sessions was counterbalanced across participants.

Each trial commenced with a white fixation cross presented for one second in the centre of a black screen, followed by a sentence, SCN or silent period. For all trials except catch trials, a silent period of 1500ms then followed, proceeded by a jittered inter-trial interval (ITI) of 1000-3000ms. For catch trials, a silent period of 500ms followed the sentence offset. The fixation cross was then replaced by a probe word that was presented for 1000ms on the screen in white 36pt bold Helvetica font. A jittered ITI also followed these sentences but ranged from 2000-3000ms to allow participants at least 3000ms from probe-onset to respond and prepare for the next trial.

In order to minimise task-related neural effects, participants were asked to attentively listen to the sentences while making every effort to understand their meaning. Participants were told that on some trials a word would appear on screen and they would have to decide whether it was either related or unrelated to the sentence they just heard. They indicated their response by pressing a button with their index or middle finger of their right hand. The pairing of “yes” and “no” response fingers was counterbalanced across subjects. In order to try to prevent participants from actively waiting for a probe to appear and ensure attention to each sentence, it was emphasised that responding to the probes would be straightforward if they listened carefully to each sentence. Participants practised the task inside the scanner before the experimental blocks. There was a higher proportion of catch-trials in the practice so that participants could familiarize themselves with the probe task.

All stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 toolbox (www.vislab.ucl.ac.uk/cogent/index.html). The visual stimuli were projected onto a screen and participants viewed them through mirrors that were attached to the head coil. MRI-compatible insert earphones (Sensimetrics, Malden, MA, USA, Model S-14) were used to deliver the auditory stimuli, providing a 20–40 dB level of attenuation. Participants wore a second set of ear protectors over the insert earphones to attenuate the scanner noise further. The experimenter checked participants could hear the sentences clearly over the scanner noise prior to the main scanning blocks.

MRI Acquisition

Participants were scanned at the Birkbeck-UCL Neuroimaging (BUCNI) Centre in London using whole-brain imaging that was performed on a Siemens Avanto 1.5T scanner. A gradient-echo EPI sequence was used to acquire the functional data (TR=3000 ms; TE=50 ms; 3×3×3mm resolution). Each run consisted of 180 volumes lasting approximately 8.5 minutes. A high-resolution anatomical scan was also acquired (T1-weighted FLASH, TR=12 ms; TE=5.6 ms; 1 mm^3 resolution) for anatomical localization purposes.

fMRI Data Analysis

The functional images were preprocessed and analysed using Statistical Parametric Mapping software (SPM8, Wellcome Department of Cognitive Neurology, London, UK). Preprocessing involved realignment, spatial normalization and smoothing (8mm Gaussian kernel) of the functional images. Three participants' data were removed because of excessive motion in the scanner in at least three runs (> 3mm). Also, one complete run of another participant was also excluded due to motion. Finally, for two participants, the final five and seven scans of one run, respectively, were excluded due to motion. Spatial normalization combined an initial affine component with subsequent non-linear warping (Friston et al., 1995) to best match the Montreal Neurological Institute's MNI-152 template. The resulting images maintained their original resolution (3x3x3mm). Two analyses were conducted with separate general linear models. The first model did not take into account the dominance conditions of the ambiguous words so that parameter estimates of the overall ambiguity effect would not be biased by the differences in sample sizes between the dominance conditions. At the first level of this model, three experimental conditions (ambiguous, unambiguous and SCN) and one "dummy" condition that included the dummy sentences and catch-trials were modelled separately. For each trial, the onset of the sentence/SCN and its duration were specified. For the catch-trials, 1.5s was added to the duration to incorporate the presentation of the visual word. The estimated movement parameters and temporal and dispersion derivatives were included as additional regressors in the model to help model structured noise in the data. At the group-level, random effects analyses were employed for two contrasts: “unambiguous versus noise” and “ambiguous versus unambiguous sentence”. For each, the corresponding contrast parameter estimates for each subject were entered into the group-level analysis, where one-sample t-tests were computed. Activations were considered significant if they reached a threshold of $p < .05$ FWE corrected at the voxel level (Worsley et al., 1996).

The second analysis was identical to the first except that dominance was also modelled. For this, the ambiguous sentences were modelled as four separate regressors corresponding to the four dominance conditions (strong biased, weak biased, balanced and dominant). For each subject, parameter estimates were obtained for three contrasts: “strong biased > unambiguous sentences”, “weak biased > unambiguous sentences” and “balanced > unambiguous”. At the group-level, these contrasts were used in a one-way repeated measures ANOVA to assess effects of dominance across the whole brain and also employed in region-of-interest (ROI) analyses, described in more detail in the Results section.

Additionally, participants’ structural images were normalised to the T1 template and, for display purposes, a mean structural image was created.

3.3 Results

Behavioural Results

Participants responded very accurately to the catch trials achieving an average accuracy of 92% (range = 79-100%), with a mean RT of 1328 ms (SD = 345), indicating that all participants were paying attention to the meaning of the sentences.

Unambiguous Sentences versus Noise

The contrast between unambiguous sentences and the low-level baseline noise condition, SCN, showed a significant broad cluster of activation in the left hemisphere centered laterally on the mid STS, extending posteriorly and anteriorly along the STS and superiorly to the anterior superior temporal gyrus (STG) (see Figure 3-2 and Table 3-3). At a lower threshold ($p < .001$ uncorrected), the left anterior temporal activation spread inferiorly into anterior middle temporal cortex. In the right hemisphere, there was a smaller significant cluster of activation that was centered in the middle STG, extending predominantly anteriorly within the STG. At the lower threshold ($p <.001$ uncorrected) it also extended posteriorly and inferiorly into right STS. There was also significant activation in the left dorsal lateral precentral gyrus. The LIFG showed activation when the threshold was lowered to $p < .001$ uncorrected, specifically within dorsal pars opercularis (peak at $x = -54$, $y = 17$, $z = 19$; z-score = 3.58).

Table 3-3: Unambiguous sentences > SCN: peak activations at p<0.05 FWE corrected. Sub-peaks that are more than 8 mm from the main peak are indented.

			Co-ordinates		
Brain region	**p(corrected)**	**Z-Score**	**x**	**y**	**z**
L STS	<.001	6.11	-54	-25	-5
L anterior STS	<.001	5.80	-57	-4	-14
L STS	.001	5.58	-60	-16	-2
L posterior STS	.011	5.08	-57	-40	7
R STG	.001	5.60	60	-10	-2
R anterior STG/STS	.030	4.88	60	-1	-11
Precentral Gyrus	.022	4.94	-48	-7	58

Note: L, Left; R, Right; STS, Superior Temporal Sulcus, STG, Superior Temporal Gyrus; MTG, Middle Temporal Gyrus.

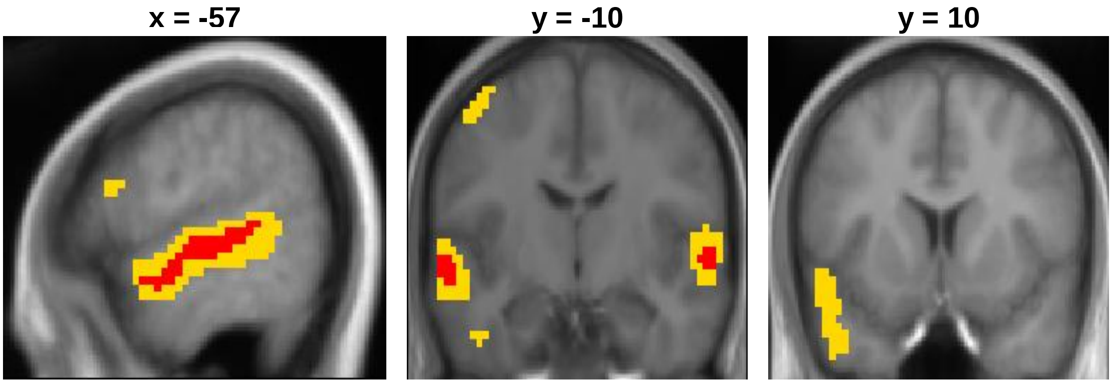

Figure 3-2: Unambiguous sentence versus SCN contrast displayed on the mean group structural image. Red represents activation significant at p<.05 FWE corrected and yellow represents activation significant at p<.001 uncorrected.

Ambiguous versus Unambiguous Sentences

Two significant clusters in the left hemisphere showed more signal for ambiguous than unambiguous sentences (see Table 3-4). One cluster was located in the LIFG, centered in pars triangularis, which extended laterally in this region. At a lower threshold ($p < .001$ uncorrected), the activation spread predominately posteriorly through pars opercularis, thereafter extending primarily dorsally in middle frontal /precentral gyrus. Activation also extended anteriorly within pars triangularis. The second cluster was located in the posterior left inferotemporal cortex (pLITC). Its peak was in the posterior occipitotemporal sulcus (OTS) but extended laterally, with a significant sub-peak in the inferior temporal gyrus (LITG). At a lower threshold ($p < .001$ uncorrected) this activation extended inferiorly into the posterior and middle portion of the fusiform gyrus, as well as superiorly through the pMTG extending along the STS. (see Figure 3-3A).

Table 3-4: Ambiguous versus unambiguous sentences: peak activations at p<0.05 FWE corrected. Sub-peaks are indented following main peak.

			Co-ordinates		
Brain region	**p(corrected)**	**Z-Score**	**x**	**y**	**z**
L OTS	.011	5.09	-45	-55	-11
L ITG	.012	5.06	-48	-58	-8
L IFG (pars triangularis)	.027	4.90	-45	32	4
L IFG (pars triangularis)	.034	4.86	-51	32	7

Note: L, Left; OTS, Occipitotemporal Sulcus; ITG, Inferior Temporal Gyrus; IFG, Inferior Frontal Gyrus.

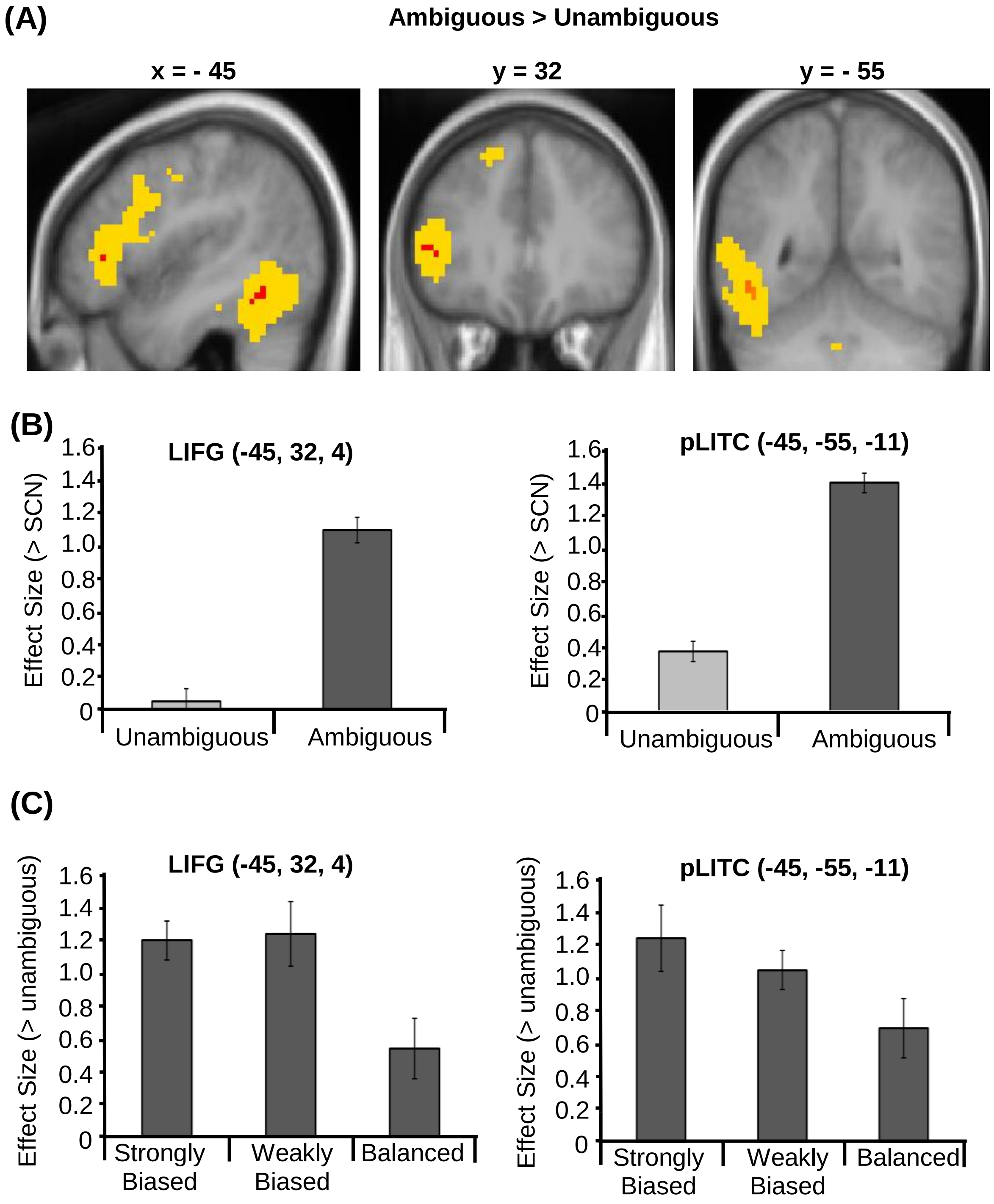

Figure 3-3: (A) Ambiguous versus Unambiguous sentence contrast displayed on the mean group structural image. Red represents activation significant at p<.05 FWE corrected and yellow represents activation significant at p<.001 uncorrected. (B) Effect sizes for the Ambiguous > SCN contrast and Unambiguous > SCN contrast averaged across the LIFG and pLITC ROIs. (C) Effect sizes for the contrasts between each dominance condition and the unambiguous sentence condition averaged across the LIFG and pLITC ROIs.

The response profiles of the frontal and temporal peak regions that showed a significant ambiguity effect were further examined with two region-of-interest (ROI) analyses. The first analysis assessed the nature of the ambiguity difference and the selectively of these regions' responses to ambiguous sentences by examining their responses to ambiguous and unambiguous sentences, separately, relative to SCN. For this, we constructed two 8mm radius spheres, one centred on the LIFG group-peak coordinate and the other centred on the pLITC group-peak, using the Marsbar toolbox (Brett, Anton, Valabregue, & Poline, 2002) and obtained averaged parameter estimates for the "Ambiguous > SCN" and the "Unambiguous > SCN" contrast for the two regions for each participant. As shown in Figure 3-3B, the ambiguity difference in both regions was, importantly, driven by increased activity for the ambiguous sentences compared to SCN rather than deactivation in the unambiguous condition. In addition, one-sample t-tests revealed that neither the LIFG nor the pLITC ROIs showed a significant response for the unambiguous sentences compared SCN (t(16) = .168, p = .869; (t(16) = 1.88, p = .253, respectively), suggesting that these regions may respond selectively for ambiguous sentences.

A second ROI analysis was conducted to assess the extent to which these two regions were affected by meaning dominance. Average parameter estimates for the strongly biased, weakly biased and balanced conditions relative to the unambiguous conditions were obtained for the LIFG and pLITC ROIs. The resulting effect sizes for each region were normalised relative to the average effect size for that ROI across all participants and all three contrasts. This was computed by dividing each subject's effect size for each contrast by the average effect size. This normalisation adjusts for differences in overall effect sizes between ROIs that may confound the magnitude of the differences found between conditions. The normalized effect sizes were entered into a 3 x 2 repeated-measures ANOVA with Dominance (strongly biased, weakly biased and balanced) as one factor and Region (LIFG, pLITC) as the other factor. The results showed a significant main effect of dominance (F(2,32) = 3.49, p = 0.042, η^2_p = .179), no significant main effect of site (F < 1) and no significant Dominance x Region interaction (F < 1), indicating that there was no reliable differences between the effect of dominance in the frontal and temporal regions. Paired t-tests between each pair of dominance conditions (averaged across region) showed that strongly biased sentences (mean = 1.23, *SD* = .80) and weakly biased sentences (mean = 1.15, *SD* = .77) produced significantly greater activation than balanced sentences (mean = .62, *SD* = .79: t(16) = 2.21, *p* = .042; t(16) = 2.19, p = .044, respectively). There was no significant difference between the strongly and weakly biased sentences (t(16) = .345, *p* = .735). See Figure 3-3C for the patterns of dominance effects for each of the ROIs.

Although the above ANOVA demonstrates that both ROIs show a similar pattern of dominance effects, it cannot demonstrate the extent to which these results are statistically reliable for each region separately. Thus, separate one-way ANOVAs were conducted on each ROI separately. The results revealed that there was a significant main of dominance for the LIFG ROI (F(2,32) = 3.56, p = 0.39), and subsequent paired t-tests confirmed that responses for the strongly and weakly biased sentences were significantly, or marginally significantly, higher than the balanced sentences (t(16) = 2.76, p = 0.14; t(16) = 1.94, p = 0.70). For the pLITC, these effects approached but did not reach significance (F(2,32) = 1.77, p = .187; t(16) = 1.51, p = .151; t(16) = 1.54, p = .143).

No significant effects of dominance were found in the whole-brain analysis (p < .05 FWE corrected).

Inter-subject Variability

Although peak co-ordinates from the group analysis identify voxels that show the most reliable effects across subjects, it is also important to assess the inter-subject variability around these peaks. For each subject we obtained the nearest local maximum (p < .05 uncorrected) to the frontal (-45, 32 4) and temporal group peaks (-45, -55, -11) for the Ambiguous > Unambiguous contrast. The location was then examined on each subject’s own structural image and identified according to sulcal landmarks. Only peaks that were within the frontal and temporal cortex were considered.

As shown in Table 3-5 and Figure 3-4, all subjects, except one, showed significant activation in close proximity to both the frontal and temporal group peaks. Only one subject did not show any significant activation around the frontal peak, with the nearest local maxima located 28mm from the peak (x = -21, y = 20, z = 13, z-score = 2.85). There was no significant difference between the two group peaks in terms of the average Euclidian distance of the local maxima (paired t-test: t(15) = 1.37, p = .190). Interestingly, the locations of these local maxima were notably more anatomically consistent (i.e., residing in the same macroanatomic region) in the frontal than in the temporal cortex. For 13 out of the 16 subjects who showed significant activation around the frontal peak, their local maxima resided in pars triangularis, with 2 additional subjects showing activation on the border between pars triangularis (PTr) and pars orbitalis (POr). In contrast, there was more anatomical variability around the temporal peak, with local maxima residing inferiorly within the ventral occipital temporal cortex areas, such as OTS and fusiform gyrus (FSG), whilst others were located more laterally within MTG/ITG.

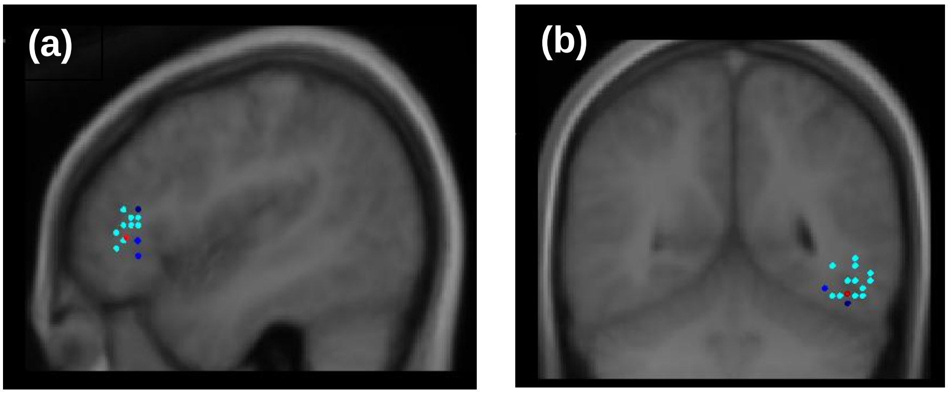

Figure 3-4: Inter-subject variability around the ambiguous vs. unambiguous contrast group peaks displayed on the group mean structural image. Red circles are the group peaks and the blue are individual subjects' peaks, with darker blue circles representing more than one subject. (a) Variability around the LIFG group peak. (b) Variability around the OTS group peak.

Table 3-5: Individual subjects' "Ambiguous > Unambiguous" local maxima nearest to the frontal and temporal group peaks.

	LIFG: -45, 32, 4						OTS:-45, -55, -11					
Subject	**x**	**y**	**z**	***Z score***	**Distance**	**Region**	**x**	**y**	**z**	***Z score***	**Distance**	**Region**
1	-36	35	7	2.20	9.9	IFS	-39	-52	-5	3.24	9.0	FSG
2	-45	32	10	2.94	6.0	PTr	-45	-49	-11	2.56	6.0	OTS
3	-45	29	-5	1.87	9.5	PTr	-48	-55	-5	2.57	6.7	MTG
4	-48	38	4	4.75	6.7	PTr	-48	-58	-11	2.06	4.2	ITG/OTS
5	-48	38	10	2.32	9.0	PTr	-45	-52	-14	2.53	4.2	OTS
6	-48	32	-5	2.62	9.5	POr/ PTr	-45	-58	-20	3.53	9.5	ITG
7	-48	26	-5	2.54	11.2	PTr	-42	-67	-14	2.39	12.7	ITG/FSG
8	-39	38	1	4.26	9.0	PTr	-39	-52	-11	2.61	6.7	OTS/ITG
9	-54	29	1	2.53	9.9	PTr	-48	-52	-14	1.81	5.2	MTG/ITG
10			n/a				-36	-52	-17	1.76	11.2	FSG
11	-51	29	-2	2.84	9.0	PTr	-57	-61	-14	3.31	13.7	MTG
12	-48	29	10	2.52	7.3	PTr	-51	-61	-20	2.83	12.4	ITG
13	-54	32	13	1.87	12.7	PTr	-48	-46	-17	2.42	11.2	OTS
14	-54	26	1	3.62	11.2	PTr	-48	-58	-11	2.13	4.2	MTG
15	-51	32	7	2.07	6.7	PTr	-36	-58	-14	3.11	9.9	FSG
16	-42	41	-2	2.02	11.2	POr./PTr	-51	-58	-11	3.63	6.7	MTG
17	-42	29	-5	1.79	9.9	PTr	-42	-61	-8	1.82	7.3	FSG
Mean	-47	32	3	2.67	9.3mm		-45	-56	-13	2.60	8.3mm	

3.4 Discussion

The results replicated the involvement of the LIFG and posterior temporal cortex in processing (temporarily) semantically ambiguous sentences seen in previous studies (Bekinschtein et al., 2011; Davis et al., 2007; Mason & Just, 2007; Rodd et al., 2005; Rodd, Longe, et al., 2010; Zempleni et al., 2007), and showed interesting effects of meaning dominance.

3.4.1 Ambiguity-Responsive Regions and Their Response Profiles

Significant activation for semantically ambiguous sentences was found in the middle portion of the LIFG and in the inferior region of the posterior temporal cortex. At a lower threshold, it was clear that these clusters belonged to larger clusters of activation that extended both anteriorly and posteriorly in the LIFG, and superiorly and ventrally within the posterior temporal cortex (see Figure 3-3A). These subthreshold activations are discussed later (section 3.4.3), mainly in terms of their implications for neural models of ambiguity resolution. Until then, discussions focus on the involvement and response profiles of the regions that reached a corrected level of significance, namely left mid-LIFG and pLITC.

The significant LIFG cluster was located in pars triangularis. This converges with the majority of published studies examining semantically ambiguous sentences, further demonstrating this to be the most consistent site of significant ambiguity-elevated peaks in the frontal cortex (Bekinschtein et al., 2011; Davis et al., 2007; Mason & Just, 2007; Rodd et al., 2005; Rodd, Longe, et al., 2010; Zempleni et al., 2007). The temporal cluster was located in pLITC, where significant peaks were found in the left posterior occipitotemporal sulcus and pLITG. This location of temporal activation is most consistent with recent studies by Rodd et al. (2012) and Bekinschtein et al. (2011), but is more inferior than other studies where activation centre around pMTG/ITG (Davis et al., 2007; Rodd et al., 2005; Zempleni et al., 2007).

The potential functions and mechanisms underlying the ambiguity-elevated responses of these two regions were further assessed by examining their response profiles to various sentence conditions (based on ROIs surrounding the frontal and temporal group peak). First the contrast between unambiguous sentences and SCN revealed that both the mid-LIFG and pLITC showed no significant response to sentences that did not contain an ambiguous word. This finding suggests these regions may not be routinely involved during speech comprehension of low-ambiguity speech and that they may therefore perform functions that are more pertinent to ambiguity resolution than sentence processing in general. These results are discussed in more detail later with regard to their implications for theories of LIFG and pLITC function.

Second, the contrast between the three types of ambiguous words (strongly biased, weakly biased and balanced) revealed that the responses of both regions were also modulated by the dominance of the ambiguous word’s meaning. Specifically, activation was greater for sentences in which the contextually appropriate meaning of the ambiguous word was its subordinate meaning (i.e., infrequently preferred by listeners) than for sentences whose meaning was one of two more equally preferred meanings. It must be noted that these dominance effects were significant for the LIFG but not for the pLITC when the ROIs were assessed separately, indicating that some caution needs to be taken when drawing conclusions about the latter region. However, the dominance effects in pLITC did not differ significantly from the LIFG and there were clear numerical effects in the same direction, which, together, provides no evidence that these regions are differently modulated by dominance.

The specific direction of modulation, where biased sentences produced greater activation than balanced sentences, suggests that these regions are especially involved when the meaning of an ambiguous word needs to be reinterpreted based on subsequent contextual information. In these sentences, disambiguating information was delayed for several words after the ambiguous word, making it likely that listeners would have committed to a single interpretation before the disambiguating information is heard (Duffy et al., 2001; Duffy et al., 1988; Rodd, Johnsrude, et al., 2010; Seidenberg et al., 1982; Swinney, 1979; Twilley & Dixon, 2000). In the case of biased sentences, this interpretation would often be the incorrect, dominant, meaning and, thus, require updating to the subordinate meaning later in the sentence. Balanced sentences would require reinterpretation less often since meaning preferences are more equal across the alternative meanings. Although no significant dominance effects were found in the whole-brain voxel-wise analysis, the increased sensitivity using an average ROI analysis is likely due to it reducing noise in the dominance-responses across voxels and subjects, which is likely to be high considering that meaning preferences are inherently variable across subjects (Rodd et al., 2013) and because the exact time-course of disambiguation is likely to vary across sentences and across individuals depending on comprehension ability (Gernsbacher & Robertson, 1995; Gernsbacher et al., 1990). Such variability may render the whole-brain analysis particularly insensitive due to the relatively modest number of trials per condition (25-33 trials).

The dominance findings converge with and extend the growing literature that report dominance effects on this neural network associated with semantic ambiguity. The results directly replicate Mason and Just's (2007) finding of greater LIFG activation for biased than balanced sentences. The trend in the temporal cortex was new. Mason and Just (2007) did not report any evidence of such a response pattern in the temporal cortex. This may be because it is a less robust effect. However, their study also did not find a significant main effect of ambiguity in this region, suggesting a general lack of sensitivity to the ambiguity response. Nevertheless, the finding that both frontal and temporal regions showed effects of dominance (albeit only numerically for the latter) is consistent with two previous studies. First, it converges with Zempleni et al.'s (2007) study in which a very similar region of the LIFG (pars opercularis/triangularis) and a slightly less inferior temporal region to that found in this study (pMTG/ITG) showed more activation for late-disambiguation sentences that biased the subordinate than dominant meaning. Second, these results also complement research on syntactically ambiguous sentences conducted by Rodd et al. (2010) in which activation in the LIFG and pMTG negatively correlated with the dominance of syntactically ambiguous sentences. Rodd et al. (2010) also examined semantically ambiguous sentences but they did not find any significant dominance correlations for these sentences. The authors hypothesised that this null finding may have been due to the position of the context. The ambiguous word was disambiguated immediately by the subsequent word in the sentence which may have limited the effect of meaning dominance as the correct meaning could be selected straight away for both biased and balanced words. The dominance effects found in the current study provide important, albeit indirect, support for this hypothesis, as the sentences employed here had a considerably longer delay between the ambiguous word and disambiguating information. Semantic reinterpretation is a key process of ambiguity resolution that has been strongly associated with the LIFG and left posterior middle/inferior temporal cortex (Rodd, Johnsrude, et al., 2012; Zempleni et al., 2007). Because dominance is an important indicator of semantic reinterpretation load, the finding that this network was modulated by this factor provides important evidence supporting its role in this process.

Interestingly, the results did not show a linear relationship between dominance and neural response, which would be expected if dominance simply reflects the likelihood of semantic reinterpretation. In fact, the two types of biased sentences patterned together: activation for strongly-biased and weakly-biased sentences was each greater than for balanced sentences but not significantly different from each other. One possible reason is that dominance scores may also index how *difficult* reinterpretation is because they also indicate how likely it is that alternative meanings are initially activated and, thus, how available the contextually appropriate meaning is when the disambiguating information is later encountered. The psycholinguistic literature does not distinguish between strongly and weakly biased words. Instead, the general consensus is that if a word has a more dominant meaning then the subordinate meaning is either not accessed or is quickly suppressed (Duffy et al., 2001; Simpson, 1994; Twilley & Dixon, 2000). In this study, the subordinate meanings of both the strongly and weakly biased sentences were much less frequent than the dominant meaning. In fact the weakly-biased sentences had an average dominance of only 21%. Thus, in both cases the subordinate meaning may not be strongly activated. In contrast, for balanced words, multiple meanings are likely to be activated and maintained for longer (Duffy et al., 2001; Simpson, 1994; Twilley & Dixon, 2000). Thus, although multiple meanings are unlikely to be maintained in parallel until disambiguating context is encountered (Duffy et al., 2001; Duffy et al., 1988; Seidenberg et al., 1982; Swinney, 1979), the non-selected meanings of balanced words may be easier to process than those of biased words because they may have a higher level of activation. Computationally, this difference in activation may arise because initially-selected, balanced meanings may not produce as much suppression as dominant meanings (Gernsbacher & St John, 2001; MacDonald et al., 1994). An alternative explanation for the similar pattern between the two types of biased words is that dominant meanings may be harder to override than those of balanced words, regardless of their level of dominance per se (Vuong & Martin, 2011). As discussed earlier, this may be because dominant meanings are more likely to be preferred, and thus integrated, more quickly than balanced meanings (Duffy et al., 2001; Gernsbacher & St John, 2001; Simpson, 1994; Twilley & Dixon, 2000). Thus, the dominance pattern suggests that this fronto-temporal network may be particularly important to integrate a less available meaning and/or suppress incorrect representations. This is highly consistent with a recent patient study demonstrating that patients with damage to the LIFG had particular difficulty in resolving subordinate-biased sentences compared to sentences with balanced ambiguous words (Vuong & Martin, 2011).

Interestingly, there was no interaction between the dominance pattern of the frontal and temporal cortex. Previous fMRI research has suggested a dissociation of ambiguity-related function between the LIFG and posterior temporal cortex, whereby the LIFG supports reinterpretation as well as initial meaning selection while the temporal cortex is only involved in reinterpretation (Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007). Similarly, neural theories of LIFG function that argue for a general conflict or semantic unification role also suggest that this region should be important for both ambiguity processes (Hagoort, 2013; Novick et al., 2005). These dominance results, however, demonstrated that the LIFG was primarily affected by reinterpretation demands. There are various reasons why it may not have shown an evident response to initial selection demands, which cannot be examined within this dataset. The simplest account is that these particular regions of the LIFG or pLITC are not involved in the initial selection between meanings, and that this process loads, instead, on other brain areas. In this current study it was not possible to distinguish whether other brain regions were important for this process as voxel-wise whole-brain analyses were found to be insensitive in detecting dominance effects, as discussed above. The alternative possibility is that the mid-LIFG, and even the pLITC, are involved in initial selection but that initial selection and reinterpretation place different demands on the processing of these regions such that processing associated with reinterpretation is more sensitive to being detected with fMRI. There are two reasons that this might be the case. First, initial selection may be a more transient process than semantic reinterpretation such that processing associated with this demand is not substantially reflected in BOLD signal which is predominately affected by local-field potentials resulting from integration of excitatory and inhibitory inputs and soma-dendritic processes occurring within a population of neurons (Logothetis, 2008). Alternatively, fMRI insensitivity to initial selection may not be due to physiological reasons per se but may be concealed by reinterpretation-related responses if they are much larger than those associated with initial selection, since both processes occur within the balanced sentences. In current models of ambiguity resolution, there is no distinction made between the relative demands arising at initial selection and semantic reinterpretation, yet, it is a possibility. Future research needs to employ experimental paradigms that can separate out the brain responses to these two processes. For example, it would be particularly useful to use methods that have a higher temporal resolution than the fMRI protocol used here as these processes occur at different times during sentence processing, such as magnetoencephalography (MEG), TMS or time-sensitive fMRI techniques (Rodd, Johnsrude, et al., 2012).

It is possible that the ambiguity-elevated responses also reflect regions responsible for anomaly detection, as oppose to reinterpretation per se, since the reinterpretation process must have been initiated by the detection of an anomaly. Consistent with this explanation are various neuroimaging studies which have found increased activation in the LIFG and/or posterior temporal cortex for semantically anomalous sentences compared to semantically congruent sentences (e.g., Baumgaertner et al., 2002; Hagoort et al., 2004; Kiehl, Laurens, & Liddle, 2002). However, the reinterpretation view seems to account for the wider range of tasks that engage these two regions across fMRI studies. In particular, these regions are engaged for various tasks that do not have any obvious anomalies, such as high-competition verb generation or picture naming tasks (e.g., Kan & Thompson-Schill, 2004a; Persson et al., 2004; Thompson-Schill et al., 1997) and cue-target semantic comparisons (e.g., Badre et al., 2005). Thus, it may actually be the case that anomalous sentences elicit activation in these regions because they initiate processes associated with reinterpretation, such as selection, inhibition and semantic retrieval, which are merely unable to be completely resolved in the anomalous cases. The fMRI data, therefore, is compatible with both anomaly detection and/or reinterpretation but the latter is more compatible with the larger neuroimaging literature.

3.4.2 Theoretical Account of Mid-LIFG

The finding of a significant ambiguity-response in the mid-LIFG region that is modulated by meaning dominance is consistent with two prominent theories of LIFG function: the conflict resolution account (Novick et al., 2009; Novick et al., 2005; Thompson-Schill et al., 1997) and the semantic unification theory (Hagoort, 2005, 2013) as both attribute functional roles to this region that are relevant for semantic reinterpretation. However, the proposed roles and their underlying mechanisms are different, suggesting alternative ways by which this region may support this process.

According to the conflict resolution account, the LIFG's role in ambiguity resolution is to resolve conflict between simultaneously active representations. In terms of semantic reinterpretation, conflict is caused when contextual information is encountered that is inconsistent with the current interpretation of an ambiguous word. In this account, the LIFG supports conflict resolution by sending top-down signals that bias information in other brain regions to support the selection of one alternative. The strong claim that this region is driven by conflict is supported by the additional finding that this region showed no significant response to unambiguous sentences over a low-level auditory baseline. Although it may be considered that unambiguous words also induce representational competition because they map onto concepts that have many alternative features which are unlikely to all be relevant to a given context , these features are not incompatible with each other and thus are unlikely to produce as much conflict as the alternative, incompatible, meanings of ambiguous words. For example, rabbits are furry, small, domesticated and shy but in the sentence “the schoolgirl was told that the rabbit was a very timid animal” only its behavioural characteristic is relevant.

The alternative, semantic unification account (Hagoort, 2005, 2013), argues that the LIFG serves to combine units of information to form larger representations of a sentence by enabling the maintenance of information online. Ambiguous words are likely to increase the demand on combinatorial processes, particularly during reinterpretation, when inconsistent contextual information needs to be integrated into a new understanding of the ambiguous word and sentence. However, the results of this study do not support one critical aspect of this theory that unification occurs for all types of sentences as there was no significant evidence that the mid-LIFG responded to unambiguous sentences. It must be noted that a more posterior cluster of the LIFG, in pars opercularis, showed a response to ambiguous and unambiguous sentences (albeit at an uncorrected threshold). Although this finding is compatible with the unification theory, this cluster was confined to posterior LIFG which is specifically attributed to syntactic and phonological unification. According to this theory, the more anterior regions should also have also shown an unambiguous response, as they support semantic unification but there was no evidence of this.

Thus, the fMRI data is not completely compatible with the claims of the unification theory, particularly with regard to the proposed role of the anterior/middle regions in semantic processing. Various other neuroimaging studies have also failed to find significant LIFG responses to sentential stimuli (Crinion et al., 2003; Rodd, Johnsrude, et al., 2012; Spitsyna et al., 2006). It is not, however, straightforward to draw conclusions from a null-result as the lack of a significant effect in the experimental condition may merely be masked by the fact that this region may also be active during the baseline conditions (J. Binder et al., 1999). Yet, patient data provide some corroborating evidence that the LIFG may not be necessary and, thus, not always involved in language comprehension. Patients with LIFG lesions have relatively preserved comprehension of words and of, albeit relatively simple, sentences (Caplan et al., 1996; Caramazza & Zurif, 1976; Novick et al., 2009; Yee et al., 2008). Thus, together the results of this study provide support for the conflict resolution account of LIFG function and, perhaps, for a revised version of the unification account in which the LIFG is viewed as mandatory for language comprehension.

3.4.3 Role of the Posterior Inferior Temporal Cortex in Ambiguity Resolution

Ambiguity-elevated activation in the pLITC was located within the middle portion of the ventral occipitotemporal cortex (vOT) (Price & Devlin, 2011). This locus was somewhat surprising as it is posterior to regions that have been associated with multimodal semantic processing (anterior fusiform/vOT: J. Binder et al., 2009; Price & Devlin, 2011) and the cluster had a more inferior distribution than that associated with sound-to-meaning mapping (MTG/ITS: Hickok & Poeppel, 2007) or semantic control (MTG: Jefferies, 2013). Instead, this region has been more generally attributed to high-level visual processing associated with either the visual form of words (Dehaene & Cohen, 2011) or with visual features of objects or meaningful stimuli more generally (Martin, 2007; Price & Devlin, 2011). In addition, it is not consistently found in auditory single word or spoken sentence studies (J. Binder et al., 2000; Davis & Gaskell, 2009; Obleser & Kotz, 2010; Spitsyna et al., 2006; Xiao et al., 2005).

However, a large body of research shows that the response of this region is strongly modulated by non-visual processes such as semantics and phonological information (Devlin et al., 2006; Song, Bu, Hu, Luo, & Liu, 2010; Twomey et al., 2011; Yoncheva, Zevin, Maurer, & McCandliss, 2010) and can be activated in the absence of visual information (e.g., Mellet, Tzourio, Denis, & Mazoyer, 1998; Price, Winterburn, et al., 2003).Thus, activation in response to ambiguity may reflect top-down accessing of visual information that may be related to orthographic representations or to visual attributes of the objects referred to in the sentence. Because two previous studies of ambiguity resolution have also found activation in this vicinity it (Bekinschtein et al., 2011; Rodd, Johnsrude, et al., 2012), it is important to understand the potential reasons why this may occur. Various reasons are considered here. Activation in this region may be driven by explicit visual imagery of the ambiguous word forms or content of the ambiguous sentence (Dehaene & Cohen, 2011; Martin, 2007). This is a possibility as the ambiguous words were the main focus of the sentence (e.g., the man thought the plant should never have been built) and the majority referred to concrete, imageable objects. The finding that this region was particularly affected by semantic reinterpretation further supports this possibility as misinterpretations may highlight the ambiguity and, thus, evoke a visual image of the word form or may evoke an image of the wrong interpretation of the sentence. This explanation is supported by a large body of research showing increased activation of visual processing areas during a variety of imagery tasks, including when participants are explicitly asked to imagine the meaning of spoken concrete nouns (D'Esposito et al., 1997; Mellet et al., 1998) and the form of visual objects (Ishai et al., 2000; O'Craven & Kanwisher, 2000) as well as when they perform a visual property verification task on pairs of words (Kan, Barsalou, Olseth Solomon, Minor, & Thompson-Schill, 2003). Alternatively, visual information may be driven automatically by inherent connections between semantic representations and perceptual attributes (Gennari et al., 2007; Kan et al., 2003; Martin, 2007; Price & Devlin, 2011; Sabsevitz, Medler, Seidenberg, & Binder, 2005). Strong support for implicit accessing of visual information comes from masked priming studies in which posterior ventral temporal responses to a written word or picture are reduced when it is preceded by a conceptually identical yet unconsciously perceived item (Kherif, Josse, & Price, 2011). This visual processing account is compatible with a recent fMRI study that showed that ambiguous verbs (“to bowl”) produced greater activation in motion-semantic areas (Gennari et al., 2007) and, thus, further suggests that different types of ambiguous words may elicit activation in different semantic areas depending on their semantic properties. Under this view, the temporal activation found in this current study may reflect the fact that the ambiguous stimuli were loading on visual semantics. However, it is not possible to test this prediction within this set of stimuli. Future research is needed to directly compare different types of ambiguous words during sentence comprehension.

If such activation reflects visual involvement, then an important question is what function does it serve? Accessing of visual information may be merely a by-product of the comprehension process (Mahon & Caramazza, 2008) or may play an essential role in resolving the competition by way of sending signals that support/inhibit the contextually appropriate/inappropriate meanings (Gennari et al., 2007). It is important to note that not all ambiguity studies have shown activation in this region. More often, activation is more lateral and superior, located more specially to the inferior and middle temporal gyri (Rodd et al., 2005; Zempleni et al., 2007). This may be due to additional sensitivity achieved in this study by having a large number of trials or may reflect differences between stimuli across studies. Future research is needed that manipulates ambiguity with imageability or varies task demands to focus attention on the visual properties of the sentence to assess whether these factors interact with ambiguity-related activation. Thus, it is clear that there are various possible explanations for ambiguity-elevated activation in the ventral temporal region primarily associated with visual processing. Further research is needed to disentangle these explanations and to understand why other areas associated with non-visual semantics are not also strongly engaged by ambiguous words.

Like accounts of LIFG function, these visual-based accounts of pITC also differ in terms of whether these imputed processes are general to sentence processing. The finding that this region showed no significant response to unambiguous sentences provides no evidence in support of a general sentence function that is, for example, argued by Martin’s (2007) theory of object knowledge. The finding, instead supports, other accounts that do not assume involvement in general sentence comprehension (Dehaene & Cohen, 2011; Price & Devlin, 2011).

While the locus of this temporal activation is most consistent with regions discussed in visual processing accounts, it must be emphasised that it is also close to regions imputed in other accounts of posterior temporal function. In particular, this region is just inferior to pMTG/ITS that is argued to support sound-meaning mapping (Hickok & Poeppel, 2007). Thus, the finding of an ambiguity effect in the broad vicinity of this region may also be considered consistent with this account, as the mapping between sound and meaning is more uncertain for ambiguous than unambiguous words. Presumably, this mapping needs re-computing when the meaning of word is not supported by contextual information (Rodd, Johnsrude, et al., 2012), which is further supported by the finding that this region was affected by reinterpretation load.

In addition, the individual subject analyses showed that for a sub-set of subjects, the ambiguity-elevated peak that was closest to the pLITC group peak was located in pMTG rather than in the inferior temporal cortex per se. This may provide some evidence for a role of this region in semantic control (Jefferies, 2013) but it is not consistent across subjects.

In summary, the results showed ambiguity-elevated activity in the vicinity of regions associated with high-level visual processing. There are various possible reasons why visual information may be elicited by ambiguous words, including that they evoke explicit imagery of the words’ referents or enhance activation of semantic features to support competition amongst alternative meanings. As these regions are close to those associated with sound-to-meaning mapping and semantic control, future research needs to assess the functional properties of the specific regions that are engaged by ambiguity to examine what kinds of information and processes they may serve in ambiguity resolution.

3.4.4 Inter-subject Variability

As this study confirms, the involvement of both frontal and temporal regions in the processing of semantically ambiguous sentences is emerging as a highly consistent finding across fMRI studies. However, these results are based on group-level analyses, which do not indicate the extent to which this reflects a network in which all components are engaged by all subjects. To investigate this, inter-subject variability was assessed in relation to the frontal and temporal two group peaks. Interestingly, all subjects (apart from one) showed ambiguity-elevated local maxima within the LIFG and posterior temporal cortex. These peaks were, on average, within 10 mm of the group peak co-ordinates. These findings validate that the group-level results reflect genuine activation patterns occurring within the majority of subjects (although it is possible that some of these activations are false positives since these effects were found at low significance thresholds).

Other interesting findings also came out of this analysis. First, the anatomical locations of the LIFG individual peaks were highly consistent, being located within pars triangularis in over 80% of subjects. This further highlights the potential importance of this particular LIFG sub-division in semantic ambiguity resolution. In contrast, the locations of the temporal peaks were more anatomically variable. While almost all subjects showed peaks in inferior temporal regions (ITG, occipitotemporal sulcus, fusiform gyrus), validating the group peak location, there was no clearly consistent anatomical field that was engaged. The nature of this variability is currently unclear. It may reflect the fact that high-level cognitive processes, such as ambiguity resolution, do not map tightly onto anatomically-defined regions (Duncan, Pattamadilok, Knierim, & Devlin, 2009; Tahmasebi et al., 2012). Alternatively, it may be more functionally relevant, perhaps reflecting the fact that individual subjects use different strategies to resolve semantic ambiguities, such as visual, phonological and/or cognitive control processes. Nevertheless, these inter-subject findings clearly show the importance of both frontal and temporal regions in processing ambiguous sentences.

3.4.5 Additional Ambiguity-responsive Regions

Inspecting the data at a lower statistical threshold revealed that ambiguity-elevated activations occurred across substantially larger clusters within the frontal and temporal cortex than that shown when applying stringent statistical threshold. The frontal cluster extended both anterior and posteriorly across the LIFG, throughout pars triangularis and pars opercularis. However, interestingly, activation was not found in its most anterior sub-division, pars orbitalis. This is particularly noteworthy as anterior LIFG has been specifically attributed to semantic processing (Badre & Wagner, 2007; Gough et al., 2005; Hagoort, 2005, 2013; Poldrack et al., 1999; Vigneau et al., 2006). It is currently unclear why these regions were not found to respond to the semantic ambiguity manipulation in this study which instead highlighted middle and posterior portions of the LIFG. This result, however, is not completely surprising as the response of anterior LIFG to semantically ambiguous sentences is the least consistent of the three sub-divisions, with some previous ambiguity studies showing the posterior pattern found in this study (Mason & Just, 2007; Rodd et al., 2005; Zempleni et al., 2007), and others reporting activation across all three sub-divisions (Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010). One potential explanation is that this region serves a specific semantic-related function that is not important to resolve all types of ambiguous sentences. For example, one current theory of the anterior LIFG is that it supports controlled semantic retrieval (Badre & Wagner, 2007). In this current study, the ambiguous sentences were disambiguated by the sentence-final word only, which meant that this word had to be strongly related to the correct meaning of the ambiguous word. As result, resolving the meaning of the ambiguous word may not have required additional controlled retrieval of semantic information that is supported by this region. Future research is clearly needed to examine the relationship between the anterior LIFG and semantic ambiguity resolution.

Another interesting observation was the notable extension of ambiguity-related activation into frontal and temporal regions that have been strongly implicated in phonological processing, namely the posterior and mid-STS as well as the posterior LIFG extending to precentral gyrus (Hagoort, 2013; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). Such activation may, thus, reflect a replaying of the heard sentence in an attempt to reanalyse the meaning of these sentences. This may be examined in variety of ways. For example, behaviourally a dual-task interference paradigm can be used to assess the effect of a concurrent phonological task on sentence comprehension (Rogalsky et al., 2008), or, a neuroimaging study could assess the neural effect of manipulating the load on phonological re-activation by manipulating the delay between an ambiguous word and the disambiguating information.

Furthermore, cognitive models of ambiguity resolution are vague in terms of the specific processes that underlie ambiguity resolution and semantic reinterpretation. It is possible that controlled semantic retrieval, phonological re-activation and/or many other processes are important. These results may provide working hypotheses for both cognitive and neural models of ambiguity resolution. Alternatively, it is possible that these less robust regions may reflect inter-subject variability in the processing of ambiguous sentences.

3.4.6 Conclusions

In conclusion, this study examined the neural responses to late-disambiguation sentences and replicated the involvement of the inferior frontal and posterior temporal cortex in ambiguity resolution. The results, furthermore, demonstrated that both regions were particularly sensitive to dominance, indicating important roles in reanalysing the meaning of a sentence during language comprehension. The next chapter examines the function of the LIFG further by using TMS to examine whether this region is causally involved in the reinterpretation process.

Chapter 4 TMS Investigation of the LIFG’s Role in Ambiguity Resolution

4.1 Introduction

An increasing number of neuroimaging studies, including the one presented in the previous chapter, support the involvement of the LIFG and posterior temporal regions in semantic ambiguity resolution (Bekinschtein et al., 2011; Mason & Just, 2007; Rodd et al., 2005; Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007). However, increases in BOLD signal do not indicate that these regions are necessary for ambiguity resolution because fMRI is a correlational measure. Blood oxygen levels are measured during the course of an event, which means that it records activity that is both necessary and incidental activity for the task. For example, detecting a visual feature of a word produces activation that is not confined to areas associated with visual processing but engages widespread activity in classical language areas (Price, Wise, & Frackowiak, 1996). Areas that process non-visual information are clearly not necessary for a non-linguistic visual-detection task. In the same manner, it is possible that some regions that show increased BOLD activity to ambiguous sentences do not play a critical role in resolving the ambiguity per se. For example, listeners may reflect on the incorrect interpretation of the ambiguous word after the correct meaning has been understood, and, thus, some activation may be a by-product of listener's awareness that a word has multiple interpretations.

4.1.1 Patient Studies

The traditional way of determining causality is by examining patients who have suffered damage to the brain region of interest. If such patients are impaired on a task, then the affected regions must be necessary for cognitive processes underlying the task. Over thirty years of patient research has focused on the function of the frontal cortex in semantic ambiguity resolution, converging on it having a causal role in this process (Bedny, Hulbert, & Thompson-Schill, 2007; Grindrod, 2012; Grindrod & Baum, 2003; Hagoort, 1993; Milberg et al., 1987; Swaab et al., 1998; Swinney et al., 1989; Vuong & Martin, 2011). Initial support came from patients with Broca's aphasia as this syndrome is associated with left frontal lobe damage. Although the specific pattern of deficits varies across studies, each study demonstrates that such patients process ambiguous words differently to neurologically healthy adults. For example, some researchers report that Broca's aphasics fail to automatically activate meanings of ambiguous words, evinced by a lack of lexical decision priming to target words presented after an ambiguous word across various priming contexts. This has been interpreted as a general deficit in automatically accessing lexical representations of words (Milberg et al., 1987). Others have found that in sentence contexts, these patients automatically activate dominant but not subordinate meanings regardless of sentence bias, which has been attributed to a slowing, rather than an absence, of lexical activation (Swinney et al., 1989). Still others report no deficit in initial lexical access, finding priming for multiple meanings, but find that patients have difficulties in using preceding context to bias activation toward the contextually appropriate meaning (Grindrod & Baum, 2003). Studies have also shown deficits in maintaining activation of the contextually appropriate meaning, as priming is found for short inter-stimulus intervals (100-500ms) between ambiguous words and targets related to these meanings but not for long ISIs (750-1250ms) (Grindrod, 2012; Grindrod & Baum, 2003; Hagoort, 1993). In contrast, another study that measured ERP to targets related to the contextually appropriate or inappropriate meanings of ambiguous sentences, showed prolonged, rather than reduced, maintenance of contextually inappropriate meanings. This suggested a deficit in using context to select the appropriate meaning that delays this process (Swaab et al., 1998). There is clearly controversy regarding the exact nature of the ambiguity-related deficits but these studies clearly converge on the broad finding that ambiguity resolution is disrupted in patients with Broca's aphasia and, thus, that the frontal lobe may be critical for this process.

Inferring a brain-behaviour relationship based on this group of patients, however, is problematic for various reasons. In all these studies patients were not specifically selected because of lesions to Broca's area. Patients were selected if they presented with the clinical syndrome that was initially discovered in patients with damage to this region (Broca, 1861) with lesion location either not being taken into account (Hagoort, 1993; Milberg et al., 1987; Swaab et al., 1998; Swinney et al., 1989) or simply being used to include patients with only left, but not right, hemisphere lesions (Grindrod, 2012; Grindrod & Baum, 2003). This is problematic because various studies have shown that this syndrome is usually not produced when lesions are restricted to the LIFG but is, instead, frequently accompanied with damage that extends beyond the LIFG and is even found in patients with non-LIFG damage, suggesting that other regions may be critical in producing this type of aphasia (Dronkers, Plaisant, Iba-Zizen, & Cabanis, 2007; Fridriksson, Bonilha, & Rorden, 2007; Mohr et al., 1978). Accordingly, in all of the aforementioned ambiguity studies, some patients either did not have damage to the LIFG or their lesion extended far beyond this region. However, a causal relationship between LIFG damage and ambiguity resolution has recently been corroborated by Vuong and Martin (2011) who showed that two patients with relatively circumscribed lesions to this region were impaired in processing the subordinate meaning of ambiguous sentences. Specifically these patients had greater difficulty than controls in reading subordinate-biased sentences compared to sentences with balanced ambiguous words or unambiguous sentences. Bedny et al. (2007) also showed that a patient with extensive damage to LIFG was impaired in suppressing the contextually inappropriate meaning of an ambiguous word. Together, such patient research suggests that LIFG is causally involved in processing the meaning of semantically ambiguous words. However, there are inconsistencies in the deficits reported across studies and, thus, it is unclear what the functional contributions are of the LIFG to this process.

In addition, all of the above studies focused on the initial processing of an ambiguous word as oppose to when the meaning of this word needs to be reanalysed based on subsequent contextual information. Neuroimaging research emphasises an important role of this region in this later process but there is no direct patient data testing this hypothesis. Some indirect support, however, comes from patient research on syntactic reinterpretation. Patients with damage to LIFG show impairments in understanding syntactic garden-path sentences, in which an early commitment to an incorrect syntactic construction is initially made, which needs to reanalysed when more of the sentence is encountered (Novick et al., 2009). This may require similar sets of cognitive processes as those needed to reanalyse the meaning of a semantically ambiguous word (MacDonald et al., 1994; Rodd, Longe, et al., 2010).

Patients with brain lesions are undoubtedly necessary for testing hypotheses regarding the causal relationship between brain regions and cognitive processes. However, such research has various experimental difficulties that affect the methodological approach that can be employed and may limit the interpretation of results. For example, the location and spread of brain damage is uncontrollable. Lesions are often large and their distribution varies widely across patients. In addition, brain damage may lead to compensatory neural changes that may lead to misinterpretations about the nature of patients' deficits. There is also the possibility that patients have a different pre-morbid ability to control subjects, which potentially confounds results when patients are compared with other populations. A number of techniques have been developed to overcome these difficulties. For example, lesion variability can be dealt with by using voxel-based lesion symptom mapping analysis, which evaluates the behavioural effects of a lesion on a voxel-by-voxel basis (Bates et al., 2003; Dronkers et al., 2004), while neural re-organisation can be assessed by measuring patients' brain activity with PET or fMRI (Meltzer, Wagage, Ryder, Solomon, & Braun, 2013; Price & Crinion, 2005; Schofield et al., 2012). However, these techniques still rely on patient data that cannot overcome the issue of pre-morbid abilities, control lesion diversity or avoid functional reorganisation confounds.

4.1.2 Transcranial Magnetic Stimulation

Transcranial Magnetic Stimulation (TMS) is an alternative neuroscientific tool that can be used to test causal relationships between brain and behaviour. One of its major advantages is that it is a non-invasive technique and thus can be used to test neurological healthy adults noninvasively, enabling it to avoid many of the limitations of patient data discussed above. TMS permits causal inference because it temporarily disrupts neural activity, producing a “virtual lesion” in the subject (Walsh & Cowey, 2000). Thus, if TMS to a brain region affects task performance then it can be concluded that that region serves a causal role in the cognitive processes underlying the task. TMS works on Faraday’s principle of electromagnetic induction, where a changing magnetic field induces an electrical current in a nearby conductive object. In TMS a rapidly changing current is delivered through a coil that produces a rapidly changing magnetic field. This generates an electric field in the brain region under the coil, affecting the underlying neuronal activity by inducing action potentials in neurons and/or changing the resting membrane potentials (T. Wagner, Rushmore, Eden, & Valero-Cabre, 2009). The consequence is that transient noise is added into the information processing that is occurring in the region, which will affect tasks that depend on these operations, usually manifesting as increases in RTs and/or errors (Devlin & Watkins, 2007; Walsh & Cowey, 2000).

TMS has several features that render it a powerful tool for examining the causal role of specific brain regions in cognitive processes such as semantic ambiguity resolution. One of the most important features is that it has both relatively high spatial and temporal resolution, which are much higher than other non-invasive stimulation techniques such as transcranial direct current stimulation (TDCS) and that are sufficiently high to allow testing dissociations between nearby brain regions and temporal dynamics.

Magnetic fields are inevitably produced under the whole coil. However, the ability to stimulate a cortical region depends on the strength of the field and the threshold that a region needs for being stimulated (Jahanshahi & Rothwell, 2000; Walsh, Pascual-

Leone, & Kosslyn, 2003). Certain configurations of TMS coils can make the field stronger at specific points on the coil, which has the effect of focalising stimulation to a relatively small region surrounding the strongest spot (Brasil-Neto et al., 1992; Ueno, Tashiro, & Harada, 1988; Wassermann, McShane, Hallett, & Cohen, 1992). Empirically, various TMS studies demonstrate that the effective spatial resolution of a figure-of-eight coil can be as precise as 0.5 – 1cm in width. For example, early studies showed that different upper limb muscles can be selectively stimulated by administering TMS to motor cortex areas that are only 0.5 - 1 cm apart (e.g., Brasil-Neto et al., 1992; Wassermann et al., 1992). In addition, TMS effects on cognitive tasks, such as semantic judgements and visual search, have been found and lost by moving the coil approximately 1 cm along the scalp (Ashbridge, Walsh, & Cowey, 1997; Gough et al., 2005). The effective focality has further been validated in a growing set of studies showing that TMS can successfully dissociate functions of nearby brain regions, including visual attention from saccades in the right superior parietal cortex (Ashbridge et al., 1997), semantic processing from phonological processing in the left inferior frontal gyrus (Gough et al., 2005) and body from face from object processing in three adjacent regions in the right extrastriate cortex (Pitcher, Charles, Devlin, Walsh, & Duchaine, 2009). Together these studies suggest that TMS can be used to target and assess the functional contribution of a relatively precise area of cortex.

TMS also has an experimentally powerful temporal resolution that can be used to target neural processing that occurs at specific time-points during a task. Some stimulation protocols such as theta-burst stimulation or repetitive low-frequency TMS induce behavioural disruption that lasts for several minutes after stimulation

(Acheson & Hagoort, 2013; Hilgetag, Theoret, & Pascual-Leone, 2001; Kosslyn et al., 1999; Pobric, Jefferies, & Lambon Ralph, 2010). However, the effects of TMS can be considerably more precise when short trains of high frequency pulses (> 1Hz) are delivered “online” during a task. For example, various studies have demonstrated that single or double pulses of TMS produce different behavioural effects at different time-points separated by as little as 10 or 40 ms (Duncan, Pattamadilok, & Devlin, 2010; Juan & Walsh, 2003; Pitcher, Goldhaber, Duchaine, Walsh, & Kanwisher, 2012; Pitcher, Walsh, Yovel, & Duchaine, 2007; Sliwinska, Khadilkar, Campbell-Ratcliffe, Quevenco, & Devlin, 2012). These results demonstrate that TMS can have a temporal resolution of tens of milliseconds. However, it must be noted that this is in stark contrast with animal research that have demonstrated that single pulses to neurons in the visual cortex of a cat produce a physiological response that can last up to seconds (Moliadze, Zhao, Eysel, & Funke, 2003). Yet, the physiological response found in animals may not equate to the durations of behaviourally-measurable disruption for a number of reasons. For example, the extent to which task performance will be disrupted depends on the extent to which the brain region can recover from the noise introduced into its processing, which may differ across individual neurons due to their different physiological properties such as size and orientation (Walsh et al., 2003). In addition, the physiological effects of brain stimulation are affected by the state of the brain region. For example, Antal, Terney, Poreisz, and Paulus (2007) showed that effects of brain stimulation are only long-lasting (i.e., for several minutes) when subjects are not using the stimulated brain region, which is the case in animal studies where the animals are anaesthetised.

The length of disruption can be flexibly adjusted by increasing the number of pulses delivered. This is particularly useful for many cognitive processes such as ambiguity resolution when it is unknown precisely when a brain region is critically involved. For example, stimulating with 3 to 5 pulses at 10Hz (i.e., for 300 – 500 ms) has been a useful protocol to induce robust behavioural impairments during a number of cognitive tasks, including visuospatial and motor attention (e.g., Rushworth, Paus, & Sipila, 2001), visual attention (e.g., Beck, Muggleton, Walsh, & Lavie, 2006), face perception (e.g., Pitcher, Walsh, & Duchaine, 2011), semantic and phonological processing (e.g.,Sliwinska et al., 2012). It is uncertain the exact temporal consequence of administering more than two TMS pulses as no time-course or animal studies have been conducted. However, insights can be gained by looking at its effects on task performance. Many TMS studies show that 3-5 pulses of 10Hz typically i) increase reaction times rather than increases errors (Devlin & Watkins, 2007; Jahanshahi & Rothwell, 2000; Walsh & Cowey, 2000), ii) slow reaction times by as little as 30-70 ms when compared to no-TMS performance (Devlin, Matthews, & Rushworth, 2003; Gough et al., 2005), and iii) can be used successfully in event-related designs such that TMS trials produce no behavioural effects on subsequent trials that do not have TMS (e.g., Devlin et al., 2003; Fuggetta, Pavone, Walsh, Kiss, & Eimer, 2006; Ganaden, Mullin, & Steeves, 2013; Gough et al., 2005). Together, these findings demonstrate that short trains of high-frequency TMS is a useful tool to assess specific time-windows that a brain region may be involved during task performance.

4.1.3 TMS Studies on Language Processing

Many studies have fruitfully used TMS to investigate the function of LIFG and many other regions in language processes (Acheson & Hagoort, 2013; Gough et al., 2005; Nixon et al., 2004; Pascual-Leone, Gates, & Dhuna, 1991; Sliwinska et al., 2012; Stewart, Meyer, Frith, & Rothwell, 2001; Watkins, Strafella, & Paus, 2003; Whitney, Kirk, et al., 2011). For example, research has demonstrated a double dissociation between the anterior and posterior LIFG in semantic and phonological processing (Gough et al., 2005) and has shown evidence for the more specific role of the anterior LIFG in semantic control (Whitney, Kirk, et al., 2011). However, despite its experimental advantages, only one study to date has used TMS to investigate semantic ambiguity resolution. Harpaz, Levkovitz, and Lavidor (2009) examined the role of left and right Wernicke's areas in processing the dominant and subordinate meanings of semantically ambiguous words by using a semantic relatedness task in which participants judged whether an ambiguous word was semantically related to a subsequent visual word (e.g., "pen-farmer"). Online TMS was employed where 5 pulses at 10Hz were delivered at the onset of the second word. Results demonstrated that stimulating the left Wernicke's area improved accuracy for dominant meaning trials whereas right hemisphere stimulation improved subordinate meaning trials. Although it is unclear why and how TMS improved performance rather than disrupting it as TMS operates by introducing interference into neural processing, the results showed that ambiguity-related processing can be successfully affected by TMS. Ambiguity studies using isolated word stimuli are undoubtedly important for understanding ambiguity processing. However, ambiguous words are usually embedded within wider contexts, such as sentences or discourses, such that processing of the ambiguous word is more naturally resolved by the surrounding context. Thus, it is particularly important to assess the neural processing of these words when they occur within sentences. Various studies have successfully investigated sentence comprehension with TMS, demonstrating involvement of the motor system in speech comprehension (Buccino et al., 2005; Tremblay, Sato, & Small, 2012), the role of LIFG in syntactic processing (Acheson & Hagoort, 2013; Sakai, Noguchi, Takeuchi, & Watanabe, 2002) and the role of the temporal cortex in idiom comprehension (Oliveri, Romero, & Papagno, 2004) and contextual integration (Franzmeier, Hutton, & Ferstl, 2012). Yet TMS has not yet been used to investigate the comprehension of semantically ambiguous sentences.

4.1.4 The Current Study

The current experiment investigated the role of LIFG in the reinterpretation of semantically ambiguous sentences by using the semantic relatedness paradigm employed in the behavioural experiments reported in Chapter 2. To reiterate, in this paradigm, participants hear late-disambiguation ambiguous sentences and unambiguous sentences, which are followed by a visual probe word that they need to decide is related or unrelated to the preceding sentence. The disambiguating information in the ambiguous sentences is delayed until the last word when reinterpretation is assumed to take place based on a wealth of ambiguity research (Duffy et al., 2001; Duffy et al., 1988; Rodd, Johnsrude, et al., 2010; Seidenberg et al., 1982; Swinney, 1979; Twilley & Dixon, 2000). The role of LIFG in semantic reinterpretation can be assessed by stimulating LIFG at the end of the sentence, when the disambiguating information is encountered, and assessing the effect of such stimulation on the semantic relatedness task. This particular behavioural task has a further functional advantage. It allows us to assess which specific functions this region serves in semantic reinterpretation because different probe conditions are sensitive to different cognitive aspects of the resolution process that are reflected in behavioural effects as demonstrated in Chapter 2. For example, RTs to related probes of ambiguous sentences (e.g., “tree” for the ambiguous sentence “the man thought the bark was going to be very damp”) relative to unambiguous sentences index activation of the contextually appropriate meaning while RTs to inappropriate probes (e.g., “sound”) compared to unrelated probes index activation of the contextually inappropriate meanings. Thus, the specific functions of LIFG on ambiguity processing can be assessed by examining the effects of stimulating the LIFG on these different probe conditions. Neural theories of ambiguity resolution do not specify which aspects of semantic reanalysis are served by the LIFG but several hypotheses can be generated. For example, if the LIFG is critical for suppressing irrelevant meanings, stimulating the LIFG should slow down responses to the inappropriate probe condition. If the LIFG is critical for using context to access the contextually appropriate meaning, LIFG stimulation should slow-down responses for related probes.

4.2 Method

Participants

26 (15 female) native English speakers took part in this experiment. Their mean age was 24 years (range: 18-47), all were right-handed, had no known reading or language impairment, nor any personal or family history of epilepsy or other

neurological condition. Participants were recruited via the UCL subject pool, board notices and student mailing lists. Participants were paid for their participation.

Stimuli

The stimuli were the same as for the behavioural experiments (Chapter 2). The only exception was that there was one less lead-in sentence presented per block (three instead of four) because, for methodological reasons, the trials were divided into more blocks (six instead of five) (see Design section).

Design

A 2 X 5 repeated-measures design was employed with TMS site (LIFG, Vertex) and probe condition (ambiguous related, ambiguous inappropriate, ambiguous unrelated, unambiguous related and unambiguous unrelated) as the two factors. As for the behavioural study, there were six versions of the stimuli, with participants performing one only version. In this TMS experiment, however, version was also divided into six blocks to enable stimulation site to be alternated across blocks. Accordingly, participants performed three blocks with LIFG stimulation and the other three with Vertex stimulation. Due to the number of unambiguous sentence items, half the blocks had an extra related probe whilst the other half had an extra unrelated probe word. Within each condition the blocks were matched on sentence dominance and on the probes' relatedness rating, frequency, log frequency, length, number of meanings and number of senses (all ps > .1). Block assignment to TMS site was counterbalanced across participants (see Procedure section).

Procedure

Participants were tested in the UCL Cognitive, Perceptual and Brain Sciences (CPB) TMS laboratory. The task was the same as the second behavioural experiment presented in Chapter 2 which employed a 300 ms sentence-probe delay. Participants heard a sentence which was followed, 300 ms later, by a visual probe word which they had to indicate was related or unrelated to the sentence. The procedure was identical to that experiment except for the following changes. TMS was delivered during the delay (see Figure 4-1) to impact processes associated with ambiguity resolution and not those related to the visual probe task. The 300 ms delay was chosen because the behavioural study demonstrated that disambiguation processing occurs during this time. Four pulses of TMS were administered, starting at the offset of the sentence-final disambiguating word. These pulses had a frequency of 10 Hz. Thus, the last pulse was delivered 300 ms post sentence offset. TMS was administered via a Magstim Rapid^2 stimulator (Magstim, Carmarthenshire, UK) and a 70 mm figure-of-eight coil. The stimulation intensity was set to 50% of the maximum stimulator output for all participants. The visual probe word was presented at the end of the TMS train, simultaneous to the last pulse.

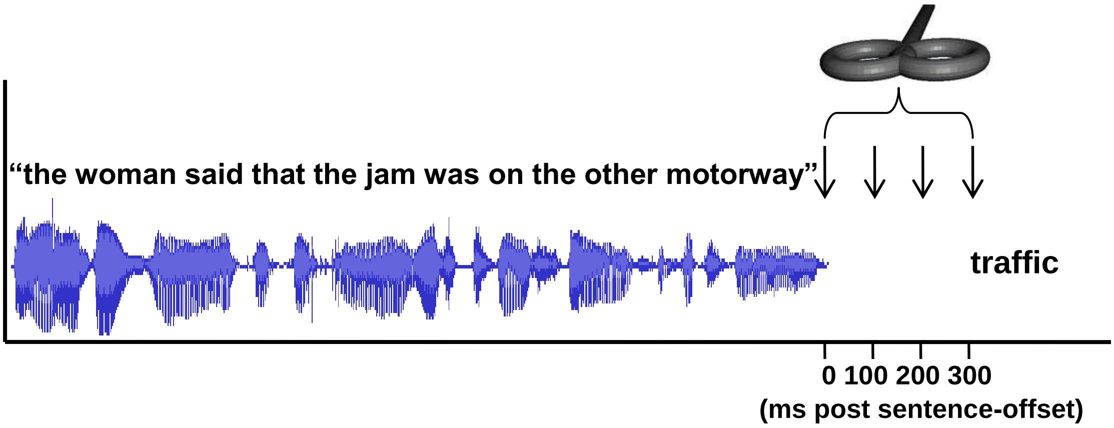

Figure 4-1: Example of a TMS trial. A spoken sentence is presented, followed by four pulses of TMS, followed by a visual probe word.

The experiment consisted of six blocks. Block order was counterbalanced using Latin squares to ensure that the same blocks did not always follow each other. Each block commenced with three lead-in items to help participants adjust to the stimulation before the experimental trials. Stimulated sites were alternated across blocks; half the participants started with LIFG stimulation whilst the other half started with Vertex stimulation. The LIFG site was centered on the MNI group peak activation co-ordinate from the fMRI experiment presented in Chapter 3 that showed a greater response to ambiguous than unambiguous sentences (x = -45, y = 32, z = 4). The LIFG and Vertex sites were localized on each participant's anatomical MRI scan. Frameless stereotaxy (Paus et al., 1997) was used to locate the TMS coil on the corresponding region of the scalp. A Polaris infrared camera (Northern Digital, Ontario, Canada) tracked the position of the participant's head and BrainSight computer software (Rogue Research, Montreal, Canada) registered the participant's head to the MRI scan. The TMS coil was positioned on an estimated location of this region and adjusted until, on the MRI scan, the estimated direction of maximum field intensity met the desired stimulation site.

Participants performed a practice block without TMS to familiarise themselves with the task. Then the practice with repeated with TMS administered to the LIFG to accustom participants to the sound and sensation of stimulation.

4.3 Results

One participant's data was excluded from all analyses due to low accuracy, achieving only 50% correct responses in the ambiguous sentence conditions. Only reaction times for correct responses were analysed. Reaction times less than 300ms were removed. Inspection of the data revealed long outliers that could be reduced by trimming to the data to exclude responses more than three standard deviations above the individual subject's mean RT. The data were analysed in the same manner as for the behavioural experiments (Chapter 2). Two main analyses were conducted. 1) A 2 x 2 x 2 ANOVA with Ambiguity (ambiguous vs. unambiguous), Relatedness (related vs. unrelated) and, now, Site (LIFG vs. Vertex) as repeated-measures factors and 2) an ANOVA comparing the ambiguous-inappropriate condition to the ambiguous-unrelated condition. Version was not included as dummy variable as there was double the number of versions as in the behavioural study rendering it less statistically useful. Items analyses were not conducted because only a small number of subjects (N = 2 – 7) contributed to the average value of each item, rendering the analyses low in statistical power and, thus, any non-significant results would be difficult to interpret.

Accuracy

The 2 x 2 x 2 (Ambiguity x Relatedness x Site) ANOVA showed a significant main effect of Ambiguity where responses were less accurate for ambiguous than unambiguous sentence probes ($F(1,24) = 24.7$, $p < .001$, $\eta^2_p = .507$; see Figure 4-2). The main effect of Relatedness was also significant with related probes less accurate than unrelated probes ($F(1,24) = 48.1$, $p < .001$, $\eta^2_p = .667$). There was also a significant Ambiguity x Relatedness interaction ($F(1,24) = 21.2$, $p < .001$, $\eta^2_p = .469$) such that accuracy was especially low for ambiguous sentence probes that were related to the sentence. There was a marginally significantly main effect of Site ($F(1,24) = 3.89$, $p = .060$, $\eta^2_p = .139$) which was, importantly, qualified by a significant Site x Relatedness interaction ($F(1,24) = 5.84$, $p = .024$, $\eta^2_p = .196$). This interaction showed that LIFG stimulation reduced accuracy for related probes only ($t(24) = 2.43$, $p = .023$). There were no other significant two or three-way interactions (all Fs < 1).

The ANOVA assessing the ambiguous-inappropriate condition revealed that these probes had significantly lower accuracy than the ambiguous-unrelated probes ($F(1,24) = 78.6$, $p < .001$, $\eta^2_p = .766$). However, there was no effect of Site nor a significant interaction between the stimulation site and probe condition (all Fs < 1).

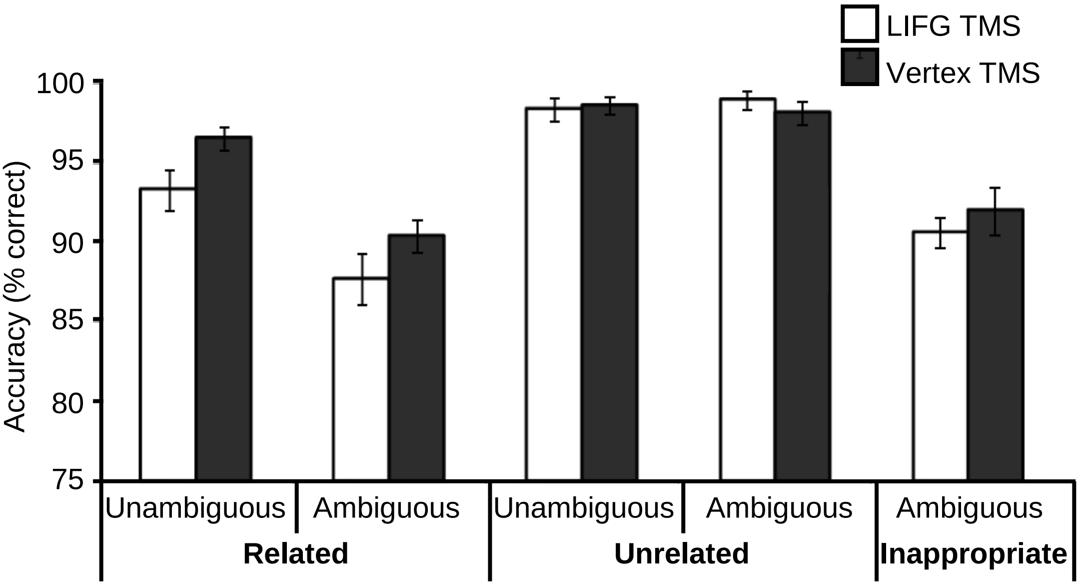

Figure 4-2: Mean accuracy (% correct) for the five probe conditions performed with LIFG (white) and Vertex (black) stimulation. Error bars show standard errors adjusted to remove between subjects variance.

Reaction times

The 2 x 2 x 2 (Ambiguity x Relatedness x Site) ANOVA showed a significant main effect of Ambiguity: responses were slower for ambiguous than unambiguous sentences (F(1,24) = 10.41, p = .004, η^2_p = .302). See Figure 4-3. The main effect of Relatedness and the Ambiguity x Relatedness interaction were not significant (F(1,24) = < 1, η^2_p = .033; F(1,24) < 1, η^2_p = .001, respectively). The main effect of Site was significant: responses were, on average slower with LIFG than Vertex stimulation (F(1,24) = 9.27, p = .006, η^2_p = .279). Neither the Ambiguity x Site interaction nor the Relatedness x Site interactions was significant (F(1,24) < 1, η^2_p = .001; F(1,24) = 1.48, p = .235, η^2_p = .058, respectively).

Importantly, the main effect of Site was qualified by a significant three-way interaction between Ambiguity, Relatedness and Site (F(1,24) = 7.64, p = .011, η^2_p = .242) The three-way interaction was evaluated in two ways. First, the effect of Site was assessed for each condition. Bonferroni-corrected paired t-tests showed that responses were only significantly slower with LIFG than Vertex stimulation for two conditions: for related probes of ambiguous sentences (t(24) = 3.15, p = .012) and for unrelated probes of unambiguous sentences (t(24) = 2.92, p = .032). The related probes of unambiguous sentences showed a relatively large numerical slow-down with LIFG stimulation, but this was not statistically reliable. This effect was half the size of that for the ambiguous-related probes and, importantly, the difference between the LIFG effect for ambiguous and unambiguous-related probes approached significance (t(24) = 2.1, p = .094).

The nature of the three-way interaction was also assessed separately for the LIFG and Vertex sites so that effects could be compared with those found in the behavioural study (see Chapter 2). Both sites produced a significant Ambiguity x Relatedness interaction (LIFG: F(1,24) = 5.07, p = .034, η^2_p = .174; Vertex: F(1,24) = 5.00, p = .035, η^2_p = .172). However, the interaction patterns were opposite to each other. Under Vertex stimulation, the ambiguity of the sentence affected responses only to unrelated probes (t(24) = 3.36, p = .003) with no significant difference between the related probes of ambiguous and unambiguous sentences (t(24) = .275, p = .786). Yet, under LIFG stimulation related probes were slower when they belonged to ambiguous sentences (t(24) = 2.81, p = .010) but there was no effect of ambiguity on the unrelated probes (t(24) = .293, p = .772).

The ANOVA analysing the ambiguous-inappropriate condition showed no significant difference between the RTs of these probes and the ambiguous-unrelated probes ($F(1,24) = < 1$, $\eta^2_p = .033$). There was also no main effect of Site ($F(1,24) = < 1$, $\eta^2_p = .002$) nor a significant interaction between the stimulation site and probe condition ($F(1,24) = < 1$, $\eta^2_p < .001$).

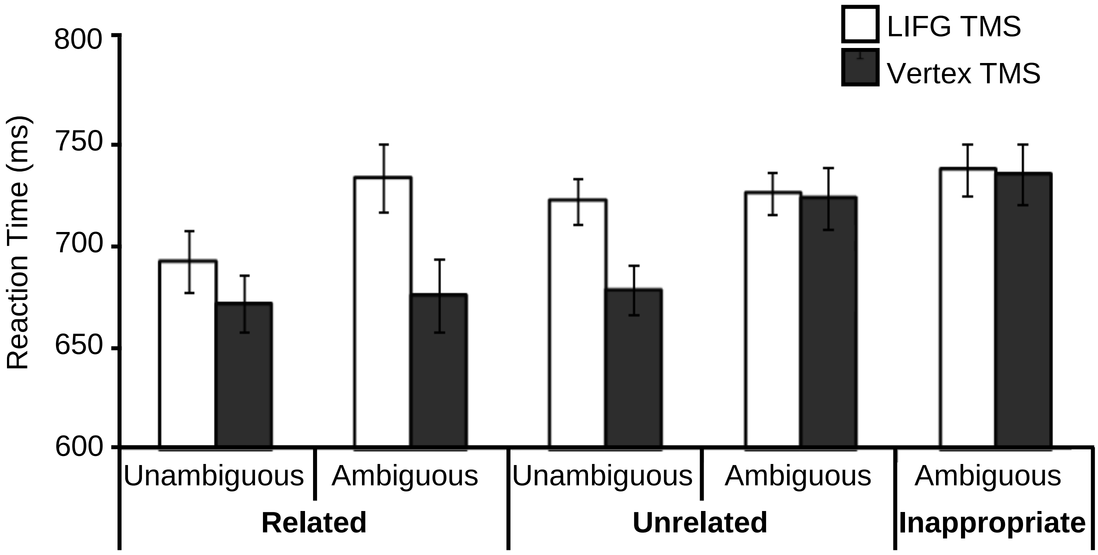

Figure 4-3: Mean reaction times for the five probe conditions performed with LIFG (white) and Vertex (black) stimulation. Error bars show standard errors adjusted to remove between subjects variance.

4.4 Discussion

This experiment investigated the causal role of the LIFG in ambiguity resolution by administering TMS to this region whilst participants listened to sentences with and without an ambiguous word. The results showed that stimulating the LIFG disrupted processing of ambiguous as well as unambiguous sentences. However, the level of disruption depended on the semantic relationship between the sentence and the probe word, which provides insights into the cognitive processes that may be supported by the LIFG.

The accuracy results were straightforward. Participants made more errors on related probes when stimulation was delivered to the LIFG compared to Vertex, regardless of the ambiguity of the sentence. Because TMS was never administered during the probe task itself, this disruption must result from interference to processes associated with sentence comprehension rather than to those associated with making the probe decision per se. The finding that both ambiguous and unambiguous sentence probes were disrupted by LIFG stimulation suggests that this region is not specific to ambiguity resolution but supports processes that are critical for comprehending even semantically straightforward sentences. Interestingly no additional reduction in accuracy was found for the ambiguous-related probes, suggesting that disrupting the LIFG functioning does not prevent disambiguation.

The reaction time data, however, showed a qualitatively different pattern of disruption. These results demonstrated a significant three-way interaction between sentence ambiguity, probe relatedness and stimulation site. Specifically the LIFG stimulation reversed the interaction pattern between sentence ambiguity and probe relatedness found under Vertex stimulation, such that ambiguity had a larger effect on the related probes than on the unrelated probes. Interestingly, this modulation resulted from LIFG stimulation increasing response times in different probes conditions of the ambiguous and unambiguous sentences. LIFG stimulation significantly slowed responses to related probes of ambiguous sentences but slowed responses to the unrelated probes of unambiguous sentences. Although only one probe condition per sentence type was significantly affected by TMS, this fits with the accuracy results in demonstrating that ambiguous as well as unambiguous sentence processing is disrupted by interfering with LIFG functioning.

The reaction time results showed some evidence that LIFG stimulation may be differently disruptive for ambiguous than unambiguous sentences. Responses for related probes of ambiguous sentences, but not those of unambiguous sentences, were significantly slower with LIFG than Vertex stimulation. Although some caution must be taken when interpreting this ambiguity difference as the direct contrast only approached significance, there was a strong trend in this direction and the numerical difference was relatively large. LIFG stimulation slowed ambiguous-related probes by on average 37ms more than the unambiguous-related probes. Related probes specifically correspond to the contextually appropriate meaning of the ambiguous word or, in the case of the unambiguous sentences, to the meaning of the unambiguous target word. Unlike the unambiguous word, the meaning of the ambiguous word critically depends on the disambiguating information that is presented at the end of the sentence when TMS is delivered. Thus, the greater slow-down for ambiguous-related probes suggests a disruption in selecting the meaning of an ambiguous word that is supported by the context.

This ambiguity disruption is highly consistent with an increasingly large set of fMRI research that show increased activation in the LIFG when participants read or listen to semantically ambiguous sentences (Bekinschtein et al., 2011; Davis et al., 2007; Mason & Just, 2007; Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007). Importantly activation is found to be greater for sentences in which semantic reinterpretation of the ambiguous word's meaning is more likely as is this case for the sentences used in this present study (Mason & Just, 2007; Rodd, Johnsrude, et al., 2012; Zempleni et al., 2007). Patient studies additionally support a causal role of the LIFG in ambiguity resolution. Patients with Broca's aphasia show disruptions in using sentence context to select the correct meaning of an ambiguous word especially when the subordinate meaning is required (e.g., Grindrod, 2012; Swaab et al., 1998; Swinney et al., 1989). Although Broca's aphasics generally have lesions that extend further than the LIFG, ambiguity resolution deficits have also been reported in patients with relatively circumscribed lesions to the LIFG including pars triangularis (Novick et al., 2009; Vuong & Martin, 2011). The TMS effect on ambiguity resolution is also in line with a recent TMS study by Acheson and Hagoort (2013) that investigated syntactically ambiguous sentences. Stimulating the LIFG affected ambiguity effects on reading times at the disambiguating region of the sentences. The nature of these ambiguity modulations was, however, somewhat unclear. On some reading measures, stimulation appeared to facilitate processing of the unambiguous sentences (e.g., produced fewer regressions into the disambiguating region than the control group) rather than impaired the ambiguous sentences (e.g., did not produce more regressions for these sentences). In contrast, the increased ambiguity effects found in this current experiment were clearly driven by the ambiguous sentence condition, providing strong evidence that the LIFG is critical for ambiguity resolution.

This finding is consistent with conflict resolution accounts of the LIFG whereby this region exerts top-down cognitive control to resolve conflict amongst representations (Novick et al., 2009; Novick et al., 2005) and select an option amongst competing alternatives (Thompson-Schill et al., 1997). In fact, selection has been particularly associated with the pars triangularis region stimulated in the current study (Badre & Wagner, 2007).

The full pattern of results, however, suggest that neither ambiguity resolution nor conflict resolution more generally is the only function of this LIFG region since stimulation also affected responses to unambiguous sentences, which induce minimal representational conflict. The effect of LIFG stimulation on unambiguous sentence processing was not completely unexpected as other theories of LIFG function argue for its importance in the semantic processing of sentences in general (e.g., Hagoort, 2005, 2013). The more surprising finding was that LIFG stimulation interacted with sentence ambiguity and probe relatedness such that it particularly disrupted the unrelated probes of unambiguous sentences. There seem to be (at least) two potential explanations for this effect. On the one hand, it is possible that the LIFG serves multiple distinct functions: one related to ambiguous sentences, such as conflict resolution (Novick et al., 2009), and another that affects comprehension of unambiguous sentences. Understanding what the latter processes might be requires understanding of the cognitive processes underlying the unrelated probe decision as this was the only unambiguous condition significantly affected by stimulation. Deciding that a probe (e.g., “rock”) is not related to a sentence (e.g., “the student had to wrap the wrist with a very old bandage”) requires a deeper consideration of the meaning of the sentence than deciding that a probe (e.g., “arm”) is related to the sentence which simply relies on noticing a semantic relationship between that word and another one in the sentence. Thus, one possible process that may be more relevant for unrelated than related probe decisions is “sentence wrap-up”. Various researchers argue that sentence wrap-up processes occur at the end of a sentence which serve to fully integrate the information presented in a sentence and resolve any representational inconsistencies that may have been encountered in order to achieve an overall coherent representation of the meaning of the sentence (Fallon, Peelle, & Wingfield, 2006; Hagoort, 2003; Just & Carpenter, 1980; Kintsch & Van Dijk, 1978; Rayner, Kambe, & Duffy, 2000). Thus, the outcome of this comprehension process seems particularly important for deciding that a probe word is not related to the sentence.

The results, however, can also be explained more parsimoniously within a single-function account of the LIFG’s role in language comprehension. Specifically, these results are compatible with Hagoort’s account of the LIFG as a general unification space that combines various sources of information to form larger complex representations of a sentence or discourse (Hagoort, 2005, 2013; Hagoort et al., 2009). All sentences require some level of unification as they contain multiple words that have different meanings. Although the meanings of unambiguous words do not depend critically on other words in the sentence as is the case for ambiguous words, the combination of their meanings is important to create a coherent understanding of a sentence. Both ambiguity resolution and sentence-wrap can be re-viewed as processes that are supported by general unification operations without needing to impute specific “conflict” or “wrap-up” operations.

Importantly, this single unification account can also suggest reasons why significant effects were not found for the other two unrelated probe conditions (ambiguous-inappropriate and ambiguous-unrelated) and neither for the unambiguous-related condition. Unambiguous-related probes (e.g., “arm”) correspond to unambiguous target words (e.g., “wrist” in the sentence “the student had to wrap the wrist with a very old bandage”). Disrupting integrational processes may not have affected the comprehension of these words because, being unambiguous, their meanings do not substantially rely on being integrated with other words in the sentence. Alternatively, critical integrational processes may still be necessary for these words but may have occurred before the end of the sentence, before TMS was delivered. This is possible as these words were presented relatively early in the sentence materials, on average, six words before the end of the sentence (e.g., “the student had to wrap the wrist with a very old bandage”), and, hence, would not have been disrupted by the stimulation.

The lack of disruption for the ambiguous-inappropriate and ambiguous-unrelated probes may be explained by a different reason related to unification processes. These probes may tap into aspects of the sentence meaning for which unification is so neurally demanding that information processing is insensitive to neuronal noise introduced by the TMS. Several reasons point to this conclusion. Under control site stimulation, responses to these two probe conditions were much slower than those of the other three conditions. The control site data further showed that the unrelated probes were significantly affected by the ambiguity of the sentence whereas related probes showed no ambiguity effect. This difference was supported by a significant interaction between sentence ambiguity and probe relatedness. Together, these findings suggest that settling on an overall understanding of a sentence, which is more important for unrelated probe decisions, is more difficult when it contains an ambiguous word and is also more difficult than merely integrating the contextually appropriate meaning of the ambiguous word. Some of this unification difficulty may arise from competition from inappropriate meanings as responses were not faster for ambiguous-inappropriate probes than ambiguous-unrelated probes, suggesting that inappropriate meanings had not been fully suppressed. My behavioural study (see Chapter 2) provides supporting evidence that suppressing inappropriate meanings is delayed relative to integrating correct meanings. One method is test this explanation is to alter the timing of TMS so that it is delivered at later stages of ambiguous sentence comprehension when processing is more advanced and, thus, according to this theory, should be more sensitive to disruption by TMS. Another method is to use offline TMS which may be less sensitive to processing difficult as it modifies the excitability of neural cells prior to a task (Thickbroom, 2007) rather than induces noise during the task. Evidently, such insensitivity to TMS disruption may also be argued to relate to conflict resolution processes (Novick et al., 2009; Novick et al., 2005), which, in the same manner, may be more neurally complex for suppressing inappropriate meanings and for building an overall representation of an ambiguous sentence. However, viewing the LIFG as important for unification processes seems to provide a parsimonious explanation for both the positive effects of disruption to the ambiguous-related and unambiguous-unrelated probe conditions as well as suggests why the remaining conditions were not significantly affected by stimulation.

Patient studies corroborate the findings of a sentence-general disruption produced by LIFG stimulation. It is well-established that patients with Broca's aphasia or lesions to the frontal cortex have impaired sentence comprehension. Although patients have greater difficulty understanding sentences that are more semantically or syntactically complex (e.g., Caplan et al., 1996; Caramazza & Zurif, 1976; Grodzinsky et al., 1999; Schwartz et al., 1980; Yee et al., 2008), some patients also show deficits for relatively simple sentences (e.g., Caplan et al., 1996; Schwartz et al., 1980) and when comprehension is successful it may result from strong lexical, semantic or pragmatic constraints (Caramazza & Zurif, 1976; Schwartz et al., 1980). Patient studies also support the importance of the LIFG in semantic processing more generally, showing that semantic priming is qualitatively different in these patients compared to neurologically-healthy participants (Grindrod, 2012; Hagoort, 1993; Swinney et al., 1989; Utman, Blumstein, & Sullivan, 2001; Yee et al., 2008).

It is interesting to note that my fMRI study (see Chapter 3) did not show increased activation in the stimulated LIFG region for this set of unambiguous sentences compared to a low-level baseline condition. This is not totally surprising as neuroimaging studies do not consistently show significant LIFG responses to unambiguous sentences (Crinion et al., 2003; Friederici, Ruschemeyer, Hahne, & Fiebach, 2003; Obleser & Kotz, 2010; Rodd, Johnsrude, et al., 2012; Spitsyna et al., 2006). Increased LIFG activation is more robustly found for sentences that are semantically or syntactically demanding, including those containing ambiguous or anomalous words (Friederici et al., 2003; Rodd, Johnsrude, et al., 2012; Rodd, Longe, et al., 2010; Zempleni et al., 2007). TMS may be a more sensitive method to examine the neural response of a region to a stimulus because it directly affects neuronal activity, inducing action potentials in cortical axons (Siebner, Hartwigsen, Kassuba, & Rothwell, 2009) whereas fMRI assesses blood oxygenation levels, which may decouple from the extent of neuronal signalling occurring in a region (Ekstrom, 2010; Logothetis et al., 2010). TMS may also be a more sensitive method because its effects are not affected by the choice of baseline in the same manner as fMRI results are. In fMRI protocols, the baseline is another task whereas in TMS protocols the baseline is the same task but performed under different stimulation parameters. The difficulty with using another task as a baseline is that it is not always clear the extent to which it may engage the region of interest. Thus, the lack of frontal activation for unambiguous sentences in fMRI studies may reflect the fact that frontal regions may be engaged during “baseline” tasks as well. Another explanation why the fMRI and TMS results may have diverged is that TMS may have stimulated two functionally-distinct regions in the LIFG. Although TMS targeted pars triangularis, which did not show an unambiguous responses in the fMRI study, the stimulation may have unintentionally also affected the posterior part of the LIFG that did show an unambiguous response (pars opercularis: -54, 17, 19). Although this region is beyond the spatial resolution of TMS, being 2.3 cm away from the pars triangularis stimulation site, it is possible that TMS affected this region indirectly. Neurophysiological studies have shown that TMS produces physiological responses in areas that are directly connected to the stimulation site (e.g., Civardi, Cantello, Asselman, & Rothwell, 2001; Mars et al., 2009). However, TMS studies on cognitive processes demonstrate that the functional consequences of a region being stimulated directly as oppose to indirectly are fundamentally different. The strongest evidence for this is that TMS can produce functional dissociations between adjacent regions that are directly connected. For example, Gough et al. (2005) showed a dissociation between the semantic and phonological functions of the anterior LIFG and posterior LIFG, respectively. Dissociations have also been demonstrated between adjacent regions of the extrastriate cortex for different perceptual processes (e.g., Pitcher et al., 2009; Sadeh et al., 2011). Thus, these dissociations suggest that the TMS disruption found for ambiguous and unambiguous sentences is unlikely to reflect the indirect stimulation of pars opercularis. Instead, the results most likely reflect disruption to the targeted pars triangularis region, highlighting a function(s) that is associated with both types of sentences. Nevertheless, future studies are needed to assess the validity of this conclusion which directly compare the effects of stimulating these two regions.

There is one caveat of this study that deserves some attention and this is the fact that TMS is also not impervious to baseline choice. The effects of LIFG stimulation are calculated relative to the control site, Vertex, data. The effects are, thus, predicted on the assumption that performance under Vertex stimulation is not different to that without TMS. Unexpectedly, the pattern of reaction times in the Vertex data showed some differences to the pattern in my behavioural experiment that employed the same task with a 300 ms sentence-probe delay (Chapter 2, Experiment 2). RTs were faster in this TMS experiment by approximately 200ms compared to the behavioural data. This is not particularly concerning as this may be simply due to nonspecific intersensory facilitation resulting from the auditory clicking sound that accompanies TMS pulses (Terao et al., 1997). More surprisingly was the different pattern of ambiguity effects. Under Vertex stimulation there was a significant interaction between sentence ambiguity and probe relatedness such that responses were slower for ambiguous than unambiguous sentences but only when the probes were unrelated to the sentence. In my behavioural study there was no significant interaction, with both related and unrelated probes showing an effect of ambiguity. In addition, here there was no significant difference between response times for ambiguous-inappropriate and ambiguous-unrelated probes whereas the behavioural study showed significantly longer responses to inappropriate probes. These findings suggest that the mere act of stimulation may have some effects on task performance that are not specific to the site that is stimulated, although it is not clear what the nature of these effects might be. As such influences may interact with the impact of LIFG stimulation, it, unfortunately, complicates the interpretation of LIFG effects and requires investigation.

In summary, the results provide novel evidence that the LIFG is necessary for resolving the meaning of ambiguous sentences as well as for comprehending sentences more generally. These disruptions may be explained by imputing multiple, ambiguity-specific and sentence-general, functions to the LIFG or by merely viewing this region as a unification space. The latter is more parsimonious and seems to fit the overall pattern of results more powerfully as it provides possible explanations for both the effects and lack of effects in all five probe conditions. However, before investigating the nature of the LIFG disruptions in more detail, it is imperative to ensure that these LIFG effects were not substantially affected by the control data. The next chapter reports a set of methodologically-driven experiments that investigates this issue.

Chapter 5 Follow-up TMS Experiments: Methodological Investigations

5.1 Experiment 6: TMS Control Condition and Order of Ambiguity Conditions

5.1.1 Introduction

The previous TMS experiment revealed several interesting effects of LIFG stimulation which suggested that this region plays a causal role in sentence comprehension. However, there were also several surprising results. Because this was the first experiment of its kind, it was particularly important to examine the extent to which such effects were influenced by theoretically uninteresting factors related to the methodological set-up. This current experiment concentrated on two potential factors, the TMS control condition and the presentation order of the ambiguity conditions, as they relate to two particularly surprising aspects of the results.

One of the most unexpected results was the RT pattern that was found in the control site (Vertex) condition because it did not show any behavioural effects of ambiguity that were found in the non-TMS behavioural study (Chapter 2, Experiment 2). This suggested that the stimulation itself may induce some non-site-specific effects on performance, making it unclear the extent to which these effects may have influenced the nature or magnitude of the LIFG effects. The second, potentially, controversial finding was the significant effect of LIFG stimulation on unambiguous sentence comprehension, since various theories and neural studies of LIFG function suggest that it is not necessary for understanding relatively straightforward sentences (Dronkers et al., 2004; Novick et al., 2009; Novick et al., 2005; Rodd, Johnsrude, et al., 2012). It is possible that some aspect of the TMS experimental set-up affected the way the unambiguous sentences were processed, making comprehension dependent on LIFG processes. For example, the unambiguous sentences may have been influenced by the ambiguous sentences because they were presented amongst each other. One possible way is as follows. The ambiguity in the ambiguous sentences is relatively salient due to the high likelihood that listeners will initially choose the wrong interpretation of the ambiguous word and need to reanalyse its meaning. As a result, participants may have experienced general uncertainty in their understanding of the sentences presented in this experiment and thus may have examined the semantic coherence of the sentence more deeply in both unambiguous and ambiguous cases.

Thus, this experiment examined the effect of 1) changing the TMS control condition and 2) presenting the unambiguous sentences separately to the ambiguous sentences. The Vertex control condition was removed. Instead, performance under LIFG stimulation was compared with that achieved when no stimulation was administered, by randomly intermixing trials where TMS was administered to the LIFG with trials without TMS. If delivering TMS during the task produces a non-site-specific effect on task performance then the pattern of responses produced without TMS should parallel those seen in the behavioural experiment data (Chapter 2, Experiment 2). If the choice of control condition has a minimal effect on the degree of LIFG disruption then we should replicate the LIFG effects found in the first experiment. That is, TMS to the LIFG should slow responses for the ambiguous-related and unambiguous-unrelated condition compared to the no-TMS condition. In addition, ambiguous and unambiguous sentences were blocked to different halves of the experiment with the unambiguous sentences presented first. If the LIFG is only recruited when there is ambiguity in the comprehension environment, LIFG stimulation should not affect the unambiguous sentence conditions. It is possible that the behavioural effects of ambiguity may be inflated by blocking ambiguity as the ambiguous sentences will not be concealed by the unambiguous sentences which may draw attention to the ambiguity manipulation. However this effect of blocking is unlikely to be substantial because the ambiguity is relatively salient in the ambiguous sentences (as mentioned above). Thus it is unlikely that the ambiguity manipulation would be noticed substantially more if the ambiguous sentences are presented without unambiguous sentences.

5.1.2 Method

Participants

10 (3 female) native English speakers took part in this experiment. The sample size is smaller than the previous experiment as this was an exploratory experiment to assess the RT effects of presenting the unambiguous sentences in a block on their own prior to the ambiguous sentences and when using a different baseline. These participants were on average 29 years old (range: 18-47), all were right-handed, had no known reading or language impairment, nor any personal or family history of epilepsy or other neurological condition. Participants were recruited from the same population as the previous experiments and were paid for their participation.

Stimuli

The stimuli were the same as in the previous TMS experiment.

Design

The design was also the same except for the following changes. Vertex stimulation was replaced with trials without TMS. These trials were randomly interleaved with trials where stimulation was administered to LIFG (i.e., there were no separate no-TMS blocks). To assign items to TMS or no-TMS trials, the items within each probe condition were divided into two lists. For each version of the experiment, the two lists of each probe condition were matched on the mean reaction time from the behavioural experiment that employed the same 300 ms probe delay (Chapter 2, Experiment 2). Such matching was conducted to ensure that any differences between TMS and no-TMS trials would not be driven by differences in baseline reaction times. Additionally, these experimental items were then divided into six blocks. In contrast to the first TMS experiment, ambiguity was blocked to different halves of the experiment such that the first three blocks contained only unambiguous sentences and the other three blocks contained only ambiguous sentences. All blocks had an even distribution of the probe conditions relevant to the ambiguity condition and an approximately equal number of TMS and no-TMS trials.

Procedure

The procedure was the same as the first TMS experiment except for the following changes. In all blocks of the experiment on half of the trials participants received stimulation to LIFG exactly as was administered in the first TMS experiment whilst on the other half of trials participants received no stimulation. TMS and no-TMS were randomly presented. In addition, the first three blocks contained only unambiguous sentences whereas the last three blocks contained the ambiguous sentences. This ambiguity manipulation was never explicitly mentioned to the participants. However, participants performed two practice blocks: one prior to the first unambiguous block and the other before the first ambiguous block. Both practices contained randomly-interleaved TMS and no-TMS trials. Before the first practice, participants were given written instructions with examples of unambiguous sentences only. After the unambiguous blocks, participants were given another set of instructions. These instructions were similar to the unambiguous sentence instructions, except that they contained examples of sentences with ambiguous words. It was important to ensure that participants treated the ambiguous inappropriate probes as unrelated to the sentences. Thus, one of the examples was of an ambiguous sentence with an ambiguous inappropriate probe and participants were informed why this probe was unrelated (i.e., that it does not fit the meaning of any of the words in that sentence). However, importantly, ambiguity was not explicitly mentioned and participants were simply told that this second set of instructions and practice block was to refresh their memory of the instructions and to ensure they still understood the task. Participants then completed the ambiguous blocks.

5.1.3 Results

The same analyses were conducted as for the previous experiment, except that the Site factor was replaced with a TMS factor (TMS vs. no-TMS).

Accuracy

The 2 x 2 x 2 (Ambiguity x Relatedness x TMS) ANOVA showed a significant main effect of Ambiguity: responses were less accurate for ambiguous than unambiguous sentence probes (F(1,9) = 5.32, p = .046, η^2_p = .372). See Figure 5-1. The main effect of Relatedness was also significant: related probes were less accurate than unrelated probes (F(1,9) = 12.8, p = .006, η^2_p = .586). The interaction between Ambiguity and Relatedness was marginally significant (F(1,9) = 4.20, p = .071, η^2_p = .318) such that accuracy was lowest for ambiguous sentence probes that were related to the sentence. The main effect of TMS was not significant (F(1,9) = .431, p = .528, η^2_p = .046) and neither were any other two-way or three way interactions (all Fs < 1). The ANOVA assessing the ambiguous-inappropriate condition revealed that these probes had significantly lower accuracy than the ambiguous-unrelated probes (F(1,9) = 13.9, p = .005, η^2_p = .608). The main effect of TMS was not significant (F(1,9) = 2.73, p = .133, η^2_p = .233). However, there was a TMS x Probe Condition interaction (F(1,9) = 7.08, p = .026, η^2_p = .440), such that accuracy for the ambiguous-inappropriate probes only was significantly better with LIFG stimulation that without TMS (t(9) = 2.56, p = .031).

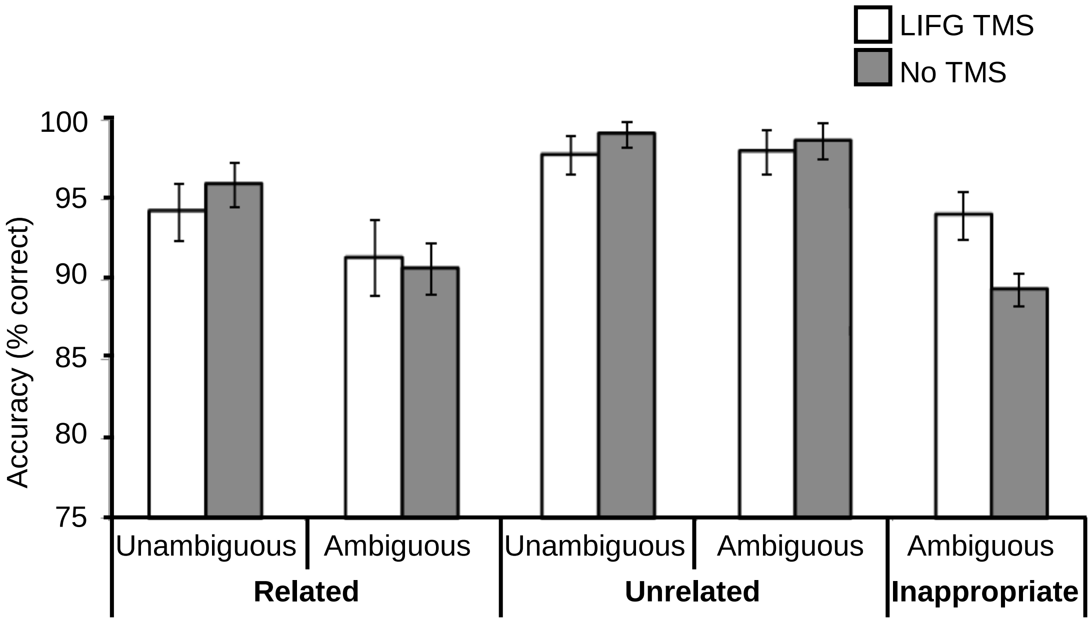

Figure 5-1: Mean accuracy (% correct) for the five probe conditions performed with LIFG stimulation (white) and with no TMS (light grey). Error bars show standard errors adjusted to remove between subject variance.

Reaction Times

It was apparent that there was substantially larger variance in the unambiguous-unrelated condition with LIFG stimulation than in the other conditions. Inspection of the data showed that this was primarily driven by one subject whose reaction time in this condition was more than 2 standard deviations above his overall mean. Removing his data did not affect the pattern of results and thus he was kept to maintain power due to the small sample size.

The 2 x 2 x 2 (Ambiguity x Relatedness x TMS) ANOVA showed a marginally significant effect of Ambiguity: responses were slower for ambiguous than unambiguous sentence probes (F(1,9) = 3.94, p = .079, η^2_p = .304). See Figure 5-2. The main effect of Relatedness was also marginally significant: responses were faster for related than unrelated probes (F(1,9) = 3.46, p = .096, η^2_p = .278). The main effect of TMS was not significant (F(1,9) < 1, η^2_p = .001) nor were any two-way or three-way interactions between the factors (all ps > .2). The ANOVA assessing the ambiguous-inappropriate condition revealed that these probes were marginally slower than the ambiguous-unrelated probes (F(1,9) = 4.42, p = .065, η^2_p = .330). Neither the main effect of TMS nor its interaction with these probe conditions was significant (both Fs > 1).

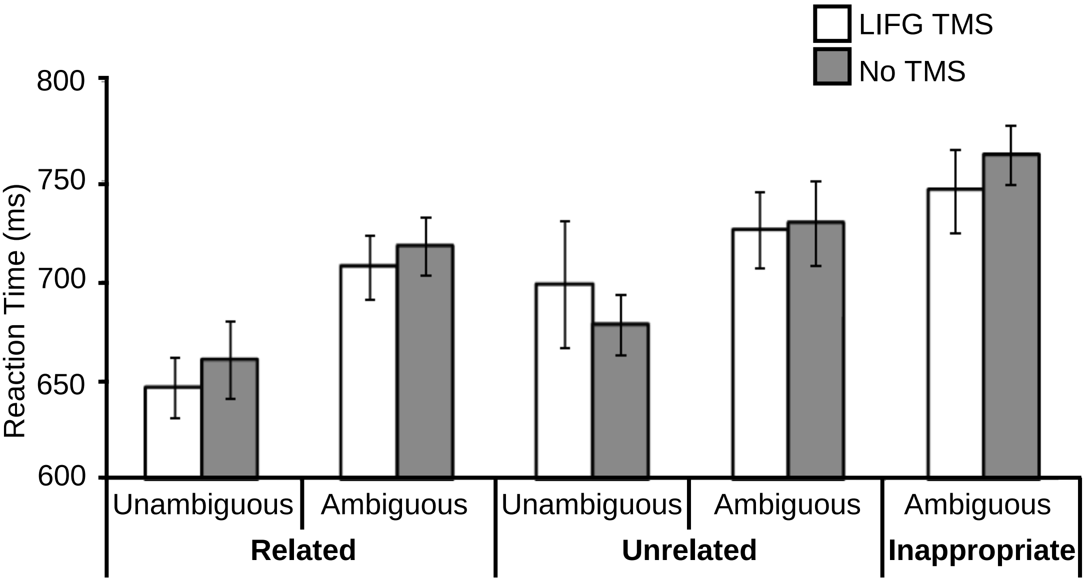

Figure 5-2: Mean reaction times for the five probe conditions performed with LIFG stimulation (white) and no TMS (light grey). Error bars show standard errors adjusted to remove between subject variance.

5.1.4 Discussion

The aim of this experiment was to examine the impact of changing the TMS control condition and blocking the order of ambiguity conditions on the pattern of behavioural effects produced under LIFG stimulation and those produced when no TMS was administered. Interestingly, the results from the no-TMS conditions replicated all the main accuracy and RT effects found in the data of the no-TMS behavioural study (Chapter 2, Experiment 2). Notably, many of these effects were marginally significant but this is most likely due to limited power resulting from the small sample size (N = 10). Ambiguity effects were evident in both the accuracy and RT data. This is particularly important for the RT data as it diverges from the pattern of RTs found in the Vertex data of the first TMS experiment (Chapter 4, Experiment 5). In contrast to the Vertex data but in line with the behavioural study, the no-TMS data showed ambiguity effects for both related and unrelated probes with no evidence of an interaction between sentence ambiguity and probe relatedness. Responses to the ambiguous-inappropriate probes were also (marginally) slower than those for the ambiguous-unrelated probes. Because these findings diverge from the Vertex data, it suggests that the act of delivering TMS during the task in Experiment 5 may have had some non-site-specific effects on task performance. Such effects may be related to the sound and sensation of stimulation that accompanies each TMS pulse. The TMS pulses were delivered during the delay between the sentence and probe word, which may have affected comprehension processes by modulating attention to the sentence task by, for example, cueing the onset of the probe word. Such effects may differently impact different aspects of sentence processing which may render certain probe conditions more sensitive to these effects.

Despite these important replications in the no-TMS conditions, the results failed to replicate any effects of LIFG stimulation found in Experiment 5. In fact, the only significant effect of TMS was found for the ambiguous-inappropriate probe condition, which surprisingly, improved with stimulation. It seems too premature to interpret this significant result because of the failure to replicate previous results. Instead, it is first important to consider reasons why LIFG stimulation was not found to affect task performance. One potential reason is the nature of the TMS control condition. It was important that the ambiguity effects were replicated in the no-TMS data as it provides strong validation of the behavioural paradigm being sensitive to the ambiguity resolution process. However, the divergence from the Vertex data suggests that using a no-TMS baseline may not be appropriate as it does not control for the non-site-specific effects of TMS that would occur when stimulation is delivered to the LIFG. This difference may have masked the effect that stimulating the LIFG has on performance, particularly if some probe conditions are more sensitive to these effects as discussed above.

An additional explanation for the lack of TMS effects concerns the level of disambiguation-related variability that may exist across items and participants. Specifically, ambiguous sentences are variable in terms of when disambiguating processes occur. The experimental sentences were created such that the last words in the sentences disambiguated the ambiguous word to its less frequent meaning. This was done to increase the likelihood that key disambiguation processes would occur upon encountering the disambiguating information, hence when TMS would be administered to the LIFG. Dominance tests confirmed that for the majority of sentences participants were more likely to select the incorrect meaning of the ambiguous word prior to the disambiguating word. However the extent to which the incorrect meaning was selected was variable across the set of sentences. Although, importantly, the correct meaning was selected on only 38% of all trials, the proportion of participants that selected the correct meaning per any individual sentence ranged from 0 – 96%. This indicates that for a small proportion of sentences participants may resolve the ambiguity before TMS is administered. Such variability suggests that TMS may not have been timed optimally for all sentences for each participant. This may have been particularly problematic in this TMS experiment because the number of trials per condition was relatively low (N = 15).

In summary, the results of this TMS experiment diverged from those found in the previous TMS experiment both in terms of effects found in the TMS control condition and regarding the effect that LIFG stimulation has on sentence comprehension. These differences may relate to the influence of non-site-specific effects of TMS on task performance or may be merely caused by high item variability that reduces sensitivity to detect effects of experimental manipulations.

5.2 Experiment 7: Larger and More Homogeneous Stimulus Set

5.2.1 Introduction

This experiment examined the degree to which high item-variability contributed to the inconsistencies between the two TMS experiments. For this, a subset of more homogeneous and better controlled ambiguous sentences was employed and the number of trials per condition was doubled by only testing the related probes of these sentences (unrelated probes of non-experimental items were included as fillers). Vertex stimulation was again adopted as the control condition in order to compare the results with those of the first TMS experiment (Chapter 4, Experiment 5).

5.2.2 Method

Participants

9 (5 female) native English speakers took part in this experiment. Like the previous follow-up experiment only a sample size was tested because this was also an exploratory experiment. Participants had a mean age of 25 years (range: 20-38), all were right-handed, had no known reading or language impairment, nor any personal or family history of epilepsy or other neurological condition. Participants were recruited from the same population as the previous experiments and were paid for their participation.

Stimuli

(i) Sentences

To increase statistical power, a subset of 60 ambiguous sentences and their matched unambiguous sentences was selected from the set of 90 sentences used in the previous TMS experiments using several criteria. First, ambiguous sentences that had $\geq$ 0.7 dominance scores at the sentence or single-word level were excluded

(N = 17) as such scores indicated a high probability that these sentences would already be disambiguated prior to the sentence-final disambiguating word (i.e., prior to stimulation). Second, sentences which had at least one probe that achieved less than an average of 70% accuracy in the behavioural study were excluded (N = 4) to ensure that all items were performed well without stimulation without any particularly difficult items. Third, ambiguous sentences whose related probe had a low mean relatedness rating (M < 4.65) were removed (N = 6) in order to statistically match the relatedness ratings of the ambiguous and unambiguous related probes. This was to ensure that any differences between the TMS effects on the ambiguous and unambiguous probe response could not be attributed to differences in relatedness ratings. Lastly, because stimulation would be administered at the offset of the sentence disambiguating word, three sentences were removed in order to better equate the frequency and number of senses of the ambiguous and unambiguous disambiguating words (N = 3). In this subset of sentences, there were no significant differences between the target words (i.e., ambiguous word or unambiguous control word) of the ambiguous and unambiguous sentences in terms of their frequency, log frequency and number of letters (paired t-tests: t(59) = .937, p = .353; t(59) = .173, p = .863; t(59) = .853, p = .397, respectively). Importantly, the ambiguous target words had a significantly higher number of meanings and senses (paired t-tests: t(59) = .853, p = .397; t(59) = .853, p = .397). There were also no significant differences between the sentence-final words of the ambiguous and unambiguous sentences in terms of their frequency, log frequency , number of letters, number of meanings and number of senses (paired t-tests: t(59) = 1.48, p = .144; t(59) = 1.75, p = .085; t(59) = 1.30, p = .199; t(59) = .830, p = .410, t(59) = 1.88, p = .065, respectively). Table

5-1 presents the descriptive statistics for characteristics of the ambiguous and unambiguous target words and the sentence-final words.

Table 5-1: Descriptive statistics (mean (SD)) for properties of the ambiguous and unambiguous target words and the sentence final words

	Frequency per million	**Log frequency**	**No. letters**	**No. meanings**	**No. of senses**
Ambiguous sentences					
Target word	54.1 (56.3)	3.58 (0.95)	4.57 (1.18)	1.88 (0.94)	9.42 (4.25)
Sentence-final word	122 (497)	3.39 (1.47)	6.93 (2.34)	1.08 (0.28)	4.43 (3.59)
Unambiguous sentences					
Target word	50.4 (45.2)	3.57 (0.93)	4.68 (1.11)	1.05 (0.22)	4.70 (2.73)
Sentence-final word	305 (1437)	3.77 (1.76)	6.60 (2.02)	1.13 (0.43)	5.73 (5.28)

Additionally, between the ambiguous and unambiguous sentences there were also no significant differences between their physical duration (paired t-tests: t(59) = .671, p = .505), number of syllables (t(59) = 1.43, p = .156) and number of words. The ambiguous sentences, however, had a significantly lower naturalness rating than the unambiguous sentences (t(59) = 3.03, p = .004). Table 5-2 presents statistics for characteristics of the whole ambiguous and unambiguous sentences.

Table 5-2: Descriptive statistics (mean (SD)) for properties of the ambiguous and unambiguous sentences

Sentence	Length in seconds	No. of syllables	No. of words	Naturalness rating
Ambiguous	2.94 (0.30)	16.2 (1.77)	12.5 (1.33)	5.47 (0.60)
Unambiguous	2.92 (0.26)	16.0 (1.74)	12.5 (1.33)	5.81 (0.60)
Fillers	2.99 (0.30)	16.7 (1.93)	12.4 (1.12)	n/a n/a

(ii) Probe words

To enhance power I also increased the number of trials per probe condition. To do this, I employed only the related probes, resulting in 30 trials per condition, which is double that used in the previous experiments. The related condition was chosen because, unlike the other probe conditions, only correct responses for related probes indicate that participants have successfully disambiguated and understood the meaning of the ambiguous and unambiguous sentences. There was no significant difference between the ambiguous and unambiguous related probes in terms of their relatedness rating (paired t-test: t(59) = 1.86, p = .068), frequency (t(59) = .212, p = .833), log frequency (t(59) = .336, p = .738), number of letters (t(59) < .001, p > .9), number of meanings (t(59) <.001, p > .9) and number of senses (t(59) = .499, p = .620). Table 5-3 presents descriptive statistics for the probe words.

Table 5-3: Descriptive statistics (mean (SD)) for properties of the probes

Probe	**Relatedness rating**	**Freq. per million**	**Log freq.**	**No. letters**	**No. meanings**	**No. senses**
Ambig.	5.69	61.8	3.38	6.15	1.08	5.50
Related	(0.61)	(82.9)	(1.29)	(1.97)	(0.33)	(5.31)
Unambig.	5.87	65.4	3.45	6.15	1.08	5.13
Related	(0.58)	(99.0)	(1.24)	(1.68)	(0.28)	(3.95)
Unrelated	n/a	76.2	3.49	5.98	1.08	5.16
Fillers	n/a	(105)	(1.44)	(1.67)	(0.35)	(4.42)

(iii) Filler sentences

Another 60 ambiguous and unambiguous sentences were employed as filler sentences to be presented with unrelated probes. These sentences were constructed in a similar way to the experimental sentences. These filler sentences had significantly fewer syllables than the set of experimental sentences (independent t-test: t(238) = 2.52, p = .012), however there were no significant differences between their number of words (t(238) = .315, p = .753) or physical duration (in seconds) (t(238) = 1.81, p = .072). See Table 5-2 for descriptive statistics for the filler sentences. An unrelated probe was chosen for each filler sentence. There were no significant differences between the unrelated and related probes' frequency (independent t-test: t(238) = .991, p = .323), log frequency (t(238) = .414, p = .679), number of letters (t(238) = .775, p = .439), number of meanings (t(238) = .197, p = .844) and number of senses (t(238) = .270, p = .787). See Table 5-3 for descriptive statistics of the unrelated probes.

Because there were two stimulation sites (LIFG and Vertex), the stimuli were divided into two lists. This division was performed for each type of sentence separately (i.e., experimental ambiguous, experimental unambiguous and filler sentences). Each pair of lists was matched on the number of syllables, number of words and physical duration (in seconds) of their sentences (independent t-tests, all ps > .1) as well as on the target words' and sentence-final words' frequency, log frequency, length (in letters), number of meanings and number of senses (independent t-tests, all ps > .075). The lists were also matched on the probe words' frequency, log frequency, length (in letters), number of meanings and number of senses (independent t-tests, all ps > .1). Each stimulation-site list was then randomly divided into three blocks (see Procedure section).

20 practice sentences were employed for use in an initial practice block and a further 18 filler sentences were employed to be presented at as lead-in items at the beginning of the experimental blocks. The sentences were constructed in the same way as the experimental sentences. Half the practice and lead-in sentences were ambiguous sentences and the other half unambiguous. All sentences were recorded by the same female speaker. For each of these sentences, one probe word was selected such that probe conditions were of roughly the same proportion as in the experimental task.

Design and Procedure

A 2 x 2 repeated-measures design was employed with TMS site (LIFG vs. Vertex) as one factor and ambiguity (ambiguous vs. unambiguous sentence) as the second factor. Accuracy and reaction times for probe judgments were the dependent variables. The two stimuli lists (see Stimuli section) were counterbalanced across stimulation sites between participants. Additionally, in contrast to the first TMS experiment, the site of stimulation was kept the same in each half of the experiment, such that in the first half participants performed three blocks of the task with stimulation to one site and in the second half they performed the remaining three blocks with stimulation to the other site. This modification was an attempt to increase consistency with which the coil was positioned on the target site. The order in which the sites were stimulated was counterbalanced across participants. The order of blocks was controlled using Latin squares to ensure that the same blocks did not always follow each other. The rest of the procedure and TMS protocol was the same as in the first TMS experiment.

5.2.3 Results

As only the related probes were of interest in this experiment, only a 2 x 2 repeated-measures ANOVA was conducted with Ambiguity (ambiguous vs. unambiguous) and Site (LIFG vs. Vertex) as the two factors.

Accuracy

This ANOVA showed a significant main effect of Ambiguity: responses were less accurate for ambiguous than unambiguous sentence probes ($F(1,8) = 7.36$, $p = .027$, $\eta^2_p = .479$). There was no significant effect of Site ($F(1,8) = .800$, $p = .397$, $\eta^2_p = .091$) and no significant Ambiguity x Site interaction ($F(1,8) = .893$, $p = .372$, $\eta^2_p = .100$). See Figure 5-3.

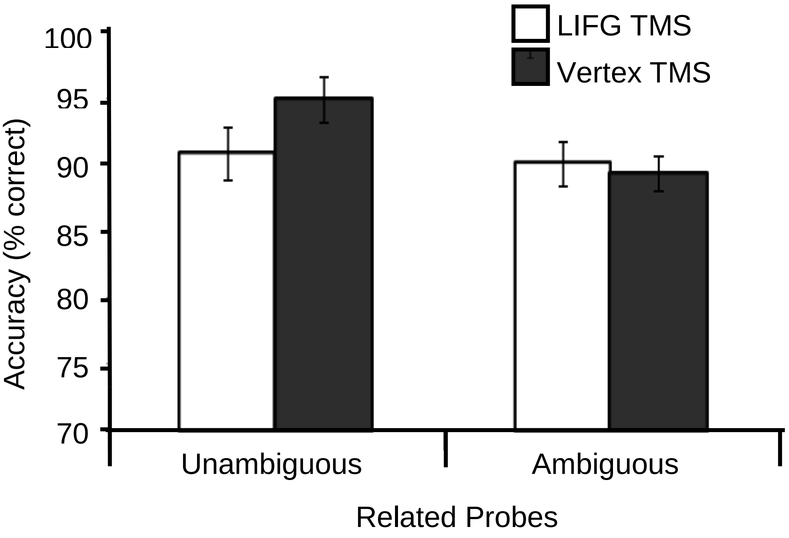

Figure 5-3: Mean accuracy (% correct) for the related probes of ambiguous and unambiguous sentences, performed with LIFG (white) and Vertex (black) stimulation. Error bars show standard errors adjusted to remove between subject variance.

Reaction Times

The ANOVA on RTs showed no significant effect of Ambiguity (F(1,8) = 1.14, p = .316, η^2_p = .125). Of note, the lack of an ambiguity effect in the Vertex condition was particularly evident as there was less than one millisecond difference between the mean RT of the ambiguous and unambiguous probes. There was also no significant effect of Site (F(1,8) = 1.07, p = .332, η^2_p = .118) or Ambiguity x Site interaction (F(1,8) < 1, η^2_p = .023).

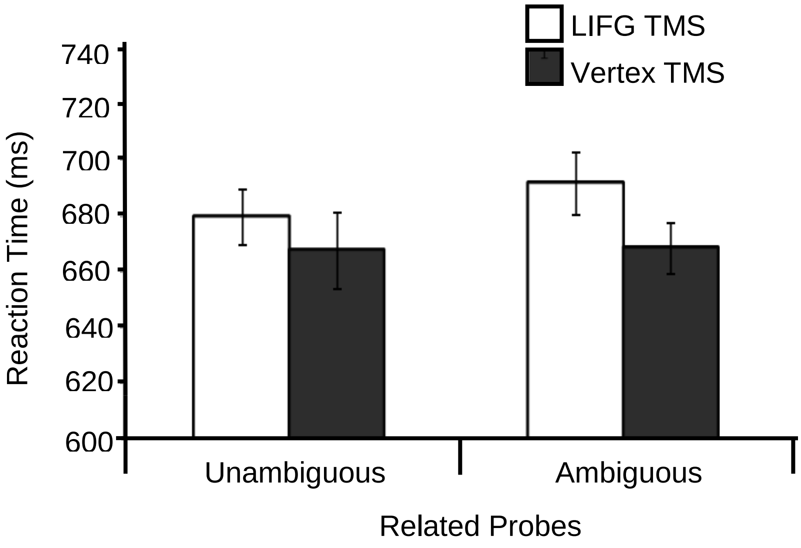

Figure 5-4: Mean reaction times for the related probes of ambiguous and unambiguous sentences, performed with LIFG (white) and Vertex (black) stimulation. Error bars show standard errors adjusted to remove between subject variance.

5.2.4 Discussion

The aim of this experiment was to assess the potential effect of high item-variability on the ability to replicate the results of the first TMS experiment. To reduce the influence of item variability, a more homogenous and better controlled set of ambiguous sentences was used with more trials per condition than that used in Experiments 5 and 6. To achieve the latter, only the related probes of each ambiguity condition were able to be included. Unrelated probes were merely included as fillers so the task would be performed properly.

Unlike the first TMS experiment, the accuracy data showed little effect of LIFG stimulation on performance. This may have resulted from removing particularly difficult items which may have reduced their sensitivity to TMS disruption, causing stimulation to slow their responses as oppose to preventing their comprehension. However, in contrast to the accuracy results, the reaction time analyses showed a highly similar pattern of results as that in the first experiment. Although there were no significant effects, which was likely due to the small sample size, the trends were strikingly apparent. First, LIFG stimulation produced slower responses, on average, than Vertex stimulation and, again, this difference was larger for the ambiguous than the unambiguous sentences. Second, the Vertex control condition showed no effect of ambiguity. In fact, the average difference between the two ambiguity conditions was less than a second. Thus, despite replication of the LIFG effects, the Vertex condition still produced a different pattern of ambiguity effects to the behavioural study, in spite of the set of better controlled ambiguous sentences and larger trial size. If Vertex stimulation is systematically affecting task performance, producing a pattern of results that is different to that seen in the behavioural experiments, then this limits the interpretability of the LIFG stimulation effects. Together, this experiment strongly suggested that the differences between the TMS results of Experiment 5 and Experiment 6 were due to the different TMS control conditions (Vertex and no-TMS, respectively) rather than being due to item variability.

5.3 Experiment 8: Vertex Stimulation versus No-TMS

5.3.1 Introduction

The preceding TMS experiments converge on the conclusion that the stimulation protocol has non-site-specific effects on task performance. This experiment directly tested this hypothesis by directly comparing participants’ performance under Vertex stimulation with their performance when no TMS is delivered.

5.3.2 Method

Participants

14 (9 female) native English speakers took part in this experiment. A relatively small sample was used as the aim was to explore whether task performance was different under Vertex stimulation than when no stimulation was administered. Participants’ mean age was 23 years (range: 18-34), all had no known reading or language impairment, nor any personal or family history of epilepsy or other neurological condition. 13 of the participants were right handed; one was left handed. Participants were recruited from the same population as the previous experiments and were paid for their participation.

Stimuli, Design and Procedure

All was the same as the previous experiment except that participants performed the task blocks either with stimulation to Vertex or with no TMS.

5.3.3 Results

The same analyses were conducted as the previous TMS experiment except that in the 2 x 2 ANOVA, the Site factor was replaced with a TMS (Vertex vs. no TMS) factor. One participant was identified as an outlier because the difference between her mean reaction time for the Vertex block and the no-TMS block (191 ms) was more than two standard deviations above the group's mean difference (M = 58 ms, SD = 55), regardless of the ambiguity condition. As the two stimulation blocks were performed in different halves of the experiment, this may indicate an unusually large effect of block order. Thus, this participant's data were excluded from all subsequent analyses.

Accuracy

This 2 x 2 (Ambiguity x TMS) ANOVA showed a significant main effect of Ambiguity: responses were less accurate for ambiguous than unambiguous sentence probes (F(1,12) = 10.2, p = .008, η^2_p = .459). There was no significant effect of TMS (F(1,12) < 1, η^2_p = .012) and no significant Ambiguity x TMS interaction (F(1,12) < 1, η^2_p = .041). See Figure 5-5.

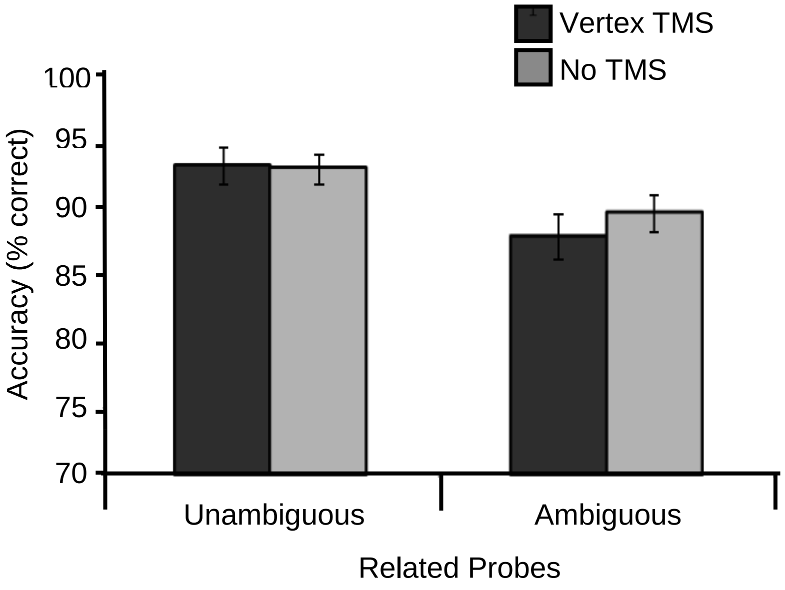

Figure 5-5: Mean accuracy (% correct) for the related probes of ambiguous and unambiguous sentences, performed with Vertex (black) and no TMS (grey) stimulation. Error bars show standard errors adjusted to remove between subject variance.

Reaction Times

The 2 x 2 (Ambiguity x TMS) ANOVA on RTs showed no significant main effect of Ambiguity (F(1,12) = 2.91, p = .114, η^2_p = .195) or TMS (F(1,12) = 2.95, p = .112, η^2_p = .197). See Figure 5-6. However, critically, there was a significant Ambiguity x TMS interaction (F(1,12) = 6.07, p = .030, η^2_p = .336) such that there was a significant ambiguity effect only when the task was performed without TMS (t(12) = 2.19, = .049). The mean ambiguity difference with Vertex stimulation was only just over 1 ms, whereas the ambiguity effect in the no TMS condition was 45 ms.

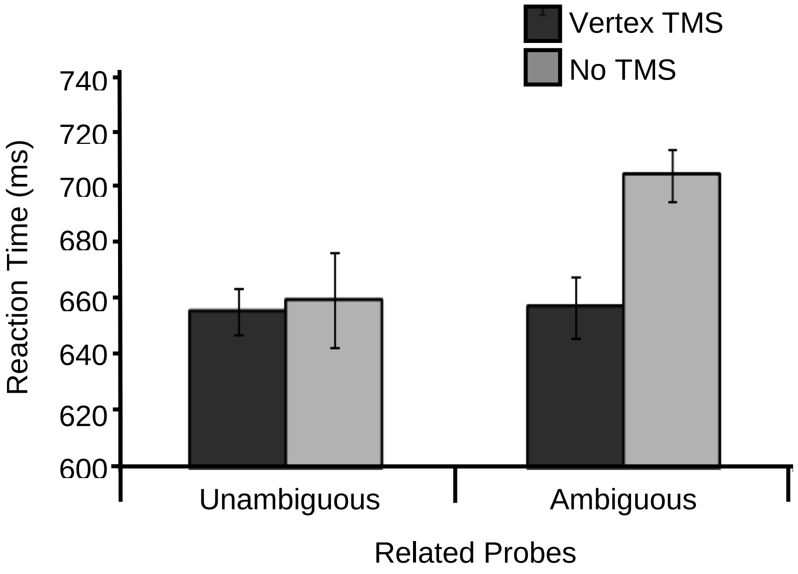

Figure 5-6: Mean reaction times for the related probes of ambiguous and unambiguous sentences, performed with Vertex (black) and no TMS (grey) stimulation. Error bars show standard errors adjusted to remove between subject variance.

5.3.4 Discussion

Confirming the previous patterns of Vertex effects, the results showed significant effects of Vertex stimulation on reaction times. Specifically, the significant interaction between ambiguity and TMS indicated that stimulating Vertex accelerated responses for ambiguous related probes but not for unambiguous ones. To be able to interpret the mechanisms behind the effects of LIFG stimulation on task performance, it is important to employ a control condition that produces effects similar to when the task is performed without stimulation.

One possible reason that Vertex stimulation may have facilitated ambiguous probe responses concerns the TMS parameters. In these experiments, TMS was administered at a frequency of 10 Hz for the whole duration of the probe delay. Specifically, pulses were separated by 100 ms, with the first applied at sentence offset and the last coinciding with presentation of the probe. This frequency produces a very rhythmic pattern of auditory clicks that may make the duration of the delay salient, highlighting the amount of time participants have to comprehend the sentence before the probe word appears. As briefly considered earlier (Experiment 5 Discussion), this may qualitatively alter the way sentences are processed during the delay period. The results demonstrated that this stimulation protocol had a specific facilitatory effect on ambiguous related probes compared to unambiguous related probes, suggesting that it particularly facilitated the processing of ambiguous sentences. It is possible that the TMS pulses acted as a “temporal cue” that substantially accelerated the disambiguation of these sentences by, for example, invoking more processing resources to discern the contextually appropriate meaning of ambiguous words such that comprehension was at an equal level for both the ambiguous and unambiguous sentences when the probe word appeared.

Although the proposed explanation is speculative, these results unquestionably demonstrate that the TMS method produces non-site-specific effects that influence the sentence-probe task in a way that appears to interact with the ambiguity manipulation. As it is unclear the extent to which LIFG functioning is modulated by these non-specific effects of TMS, the TMS parameters need to be changed to eliminate potential confounds arising from these effects.

5.4 Experiment 9: 20Hz Stimulation to Vertex vs. No-TMS

5.4.1 Introduction

This experiment sought to examine the non-specific effects of a different set of TMS parameters on the sentence-probe task. The aim was to employ a set of parameters that reduces the distinct rhythmic pattern of the pulses produced with 10 Hz and that also reduces its ability to cue the length of the probe delay. Thus, frequency was increased to 20 Hz which reduces the rhythmic pattern as the pulses are much closer in time. The same number of pulses (four) was delivered starting from the sentence offset so that the same amount of auditory and tactile stimulation occurred as the previous experiments. This has the additional effect of making the last pulse occur 150ms into the delay, which does not coincide with, and thus should not "cue", the presentation of the probe word that is still presented 300ms post sentence-offset. See Figure 5-7. If the pattern of Vertex data found in the previous TMS experiments (Experiment 5, 7, 8) were due to attentional cueing effects of 10 Hz stimulation, then these effects should be reduced with 20 Hz stimulation.

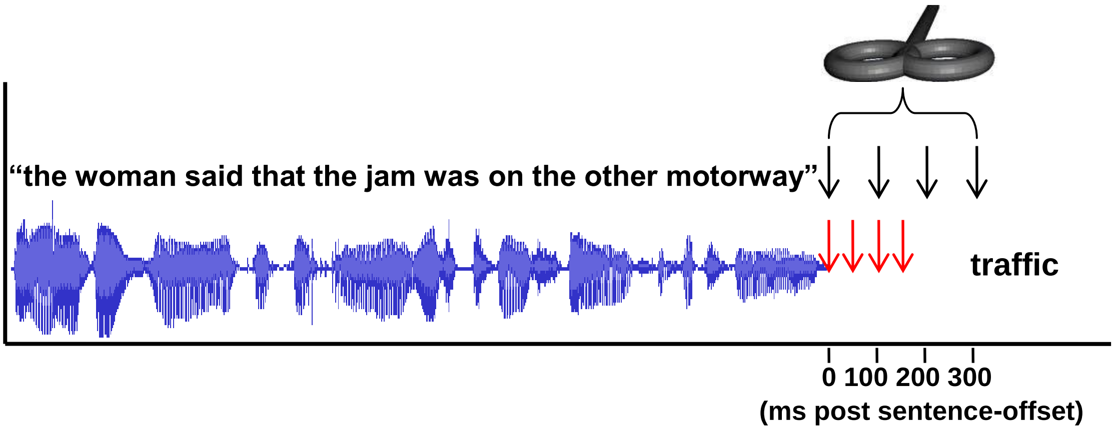

Figure 5-7: Example of a TMS trial. A spoken sentence is presented, followed by four pulses of TMS, followed by a visual probe word. The red arrows represent 20 Hz stimulation and the black arrows represent 10 Hz stimulation for comparison.

5.4.2 Method

Participants

The same sample size was tested as the previous TMS experiment: 14 (9 female) native English speakers. Their mean age was 26 years (range: 19-44), all were right-handed, had no known reading or language impairment, nor any personal or family history of epilepsy or other neurological condition. Participants were recruited from the same population as the previous experiments and were paid for their participation.

Stimuli, Design and Procedure

All was the same as the first TMS experiment (Chapter 4, Experiment 5) except that the frequency of TMS was changed to 20 Hz, such that 4 pulses of TMS were administered from sentence offset, each separated by 50ms.

5.4.3 Results

Results were analysed in the same way as the previous TMS experiment. As in the previous experiment's data, one participant was identified as an outlier because the difference between her mean reaction time for the Vertex block and the no-TMS block (164 ms) was more than two standard deviations above the group's mean difference (M = 47 ms, SD = 49), regardless of the ambiguity condition. As the two stimulation blocks were performed in different halves of the experiment, this may indicate an unusually large effect of block order. Thus, this participant's data were excluded from all subsequent analyses.

Accuracy

Accuracy was high across the four conditions. See Figure 5-8. The 2 x 2 (Ambiguity x TMS) ANOVA showed no significant effect of Ambiguity, TMS or interaction (F(1,12) < 1, η^2_p = .028; F(1,12) = 1.39, p = 261, η^2_p = .104; F(1,12) < 1, η^2_p = .053, respectively).

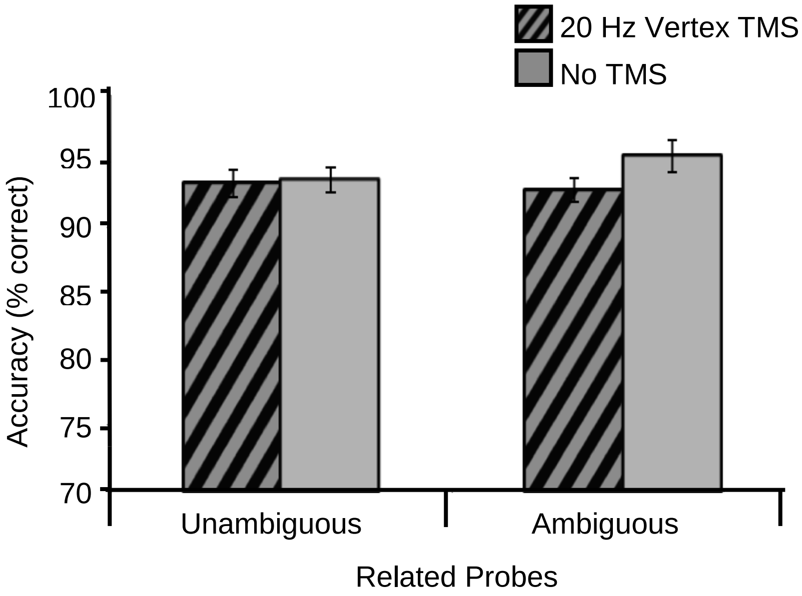

Figure 5-8: Mean accuracy (% correct) for the related probes of ambiguous and unambiguous sentences, performed with 20Hz Vertex TMS (striped) and no TMS (grey) stimulation. Error bars show standard errors adjusted to remove between subject variance.

Reaction Times

The 2 x 2 (Ambiguity x TMS) ANOVA on RTs showed no significant main effect of Ambiguity, although there was a trend for responses to be slower for ambiguous than unambiguous sentence probes ($F(1,12) = 2.86$, $p = .117$, $\eta^2_p = .192$). See Figure 5-9. There was also a significant main effect of TMS such that responses were faster with Vertex stimulation than without TMS ($F(1,12) = 5.91$, $p = .032$, $\eta^2_p = .330$). However, critically, TMS did not significantly interact with Ambiguity ($F(1,12) < 1$, $\eta^2_p = .056$) such that *both* TMS conditions showed an ambiguity effect that approached significance. The mean ambiguity difference was 25 ms under Vertex stimulation and 38 ms without TMS (t(12) = 1.66, p = .123; t(12) = 1.56, p = .144, respectively).

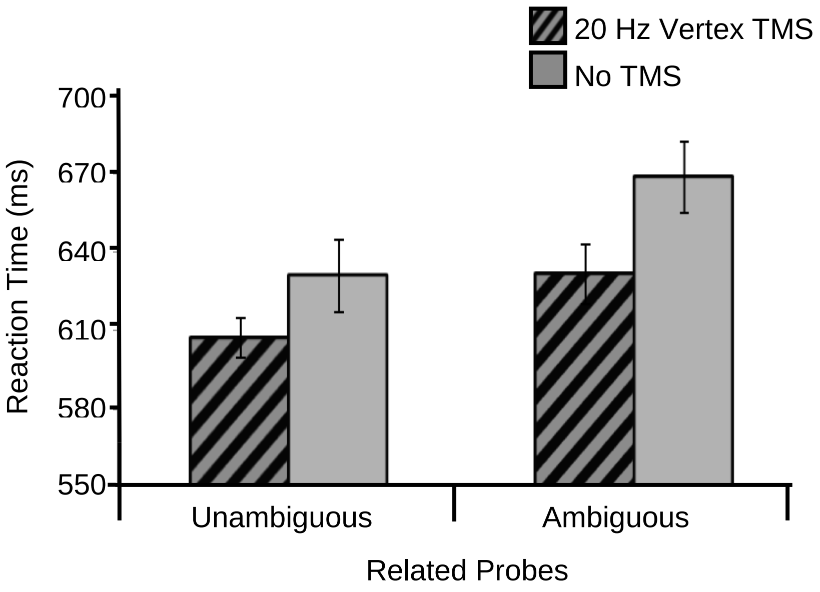

Figure 5-9: Mean reaction times for the related probes of ambiguous and unambiguous sentences, performed with 20 Hz Vertex TMS (striped) and no TMS (grey) stimulation. Error bars show standard errors adjusted to remove between subject variance.

5.4.4 Discussion

The reaction times analyses revealed a significant main effect of TMS but this is not of great concern as it is likely due to simple intersensory facilitation that is induced by the auditory clicks that accompany TMS pulses (Terao et al., 1997). The important finding was that this general TMS effect did not interact with sentence ambiguity, indicating that stimulation did not differently affect ambiguous and unambiguous sentences. In addition, there was evidence of ambiguity effects in both the TMS condition and the no-TMS condition. These results are clearly different to those produced by 10 Hz stimulation and, reassuringly, are highly consistent with the results of the behavioural study. Thus, this protocol was subsequently used to test the effect of LIFG stimulation in the next TMS experiment.

5.5 Experiment 10: Re-examining the Effects of LIFG Stimulation on Sentence Comprehension

5.5.1 Introduction

The aim of this experiment was to re-examine the effects of LIFG stimulation on sentence comprehension by employing the new TMS parameters (4 pulses at 20 Hz). In contrast to 10 Hz TMS, these parameters produce similar behavioural effects of ambiguity resolution when administered to a control site as that produced without TMS. This, importantly, eliminates potential confounds on the interpretation of effects found for LIFG stimulation. If the LIFG plays a necessary role in ambiguity resolution then LIFG stimulation should affect responses to ambiguous sentence probes relative to Vertex stimulation. If the LIFG is also involved in sentence comprehension more generally, then LIFG stimulation should also affect responses to unambiguous sentence probes. To maximise power only the related probes of the ambiguous and unambiguous sentences were examined as in Experiments 6 – 8.

5.5.2 Method

Participants

21 (15 female) native English speakers took part in this experiment. Their mean age was 21 years (range: 18-33), all were right-handed, had no known reading or

language impairment, nor any personal or family history of epilepsy or other neurological condition. Participants were recruited from the same population as the previous experiments.

Stimuli, Design and Procedure

All was the same as TMS Experiment 3 except that the frequency of TMS was changed to 20 Hz, such that 4 pulses of TMS were administered from sentence offset, each separated by 50 ms. In addition, the stimulation sites were alternated across blocks, as in Experiment 1, in order to eliminate general effects of block order.

5.5.3 Results

One participant's data was excluded from all analyses due to low accuracy (mean accuracy = 58%). By-items analyses (F_2) were conducted because a reasonable number of observations contributed to each item's average due to the experiment having only versions. One ambiguous item (ambiguous word: 'pack', related probe: 'herd') was also excluded from these analyses due to low accuracy (30%), which may result in unreliable mean estimates. Its unambiguous counterpart was also removed.

Accuracy

The 2 x 2 (Ambiguity x Site) ANOVA showed a significant main effect of Ambiguity in the by-subjects analysis such that accuracy was lower for ambiguous than unambiguous sentence probes ($F_1(1,19) = 5.28$, $p = .033$, $\eta^2_p = .218$; $F_2(1,58) =$

1.89, $p = .174$, $\eta^2_p = .032$). There was, however, no significant effect of Site and no Site x Ambiguity interaction (both *Fs* < 1). See Figure 5-10.

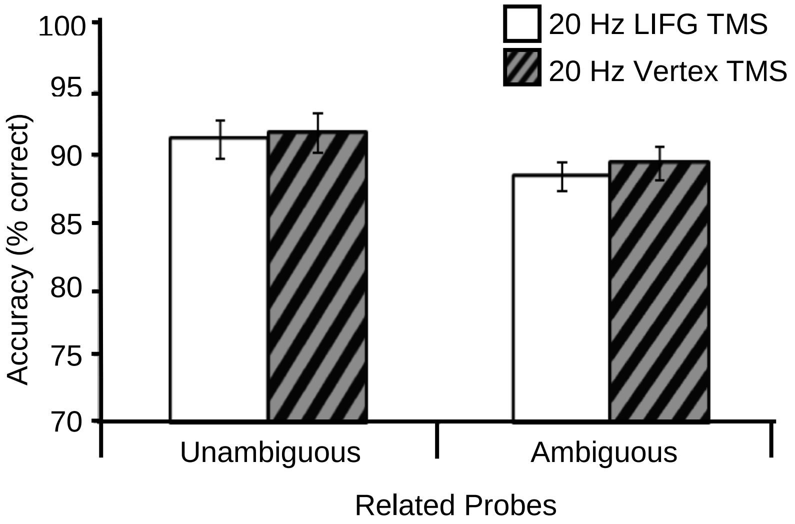

Figure 5-10: Mean accuracy (% correct) for the related probes of ambiguous and unambiguous sentences, performed with 20 Hz LIFG TMS (white) and 20 Hz Vertex TMS (striped) stimulation. Error bars show standard error adjusted to remove between subject variance.

Reaction Times

The 2 x 2 (Ambiguity x Site) ANOVA on RTs showed a marginally significant main effect of Ambiguity: responses were slower for ambiguous than unambiguous sentence probes ($F_1(1,19) = 4.01$, $p = .060$, $\eta^2_p = .174$; $F_2(1,19) = 2.66$, $p = .108$, $\eta^2_p = .044$). See Figure 5-11. Critically, the main effect of TMS was significant in both the by-subjects and by-items analyses: responses were slower under LIFG than

Vertex stimulation ($F_1(1,19) = 4.42$, $p = .049$; $\eta^2_p = .189$; $F_2(1,19) = 10.14$, $p = .002$, $\eta^2_p = .149$). The Ambiguity x TMS interaction was not significant (all Fs < 1).

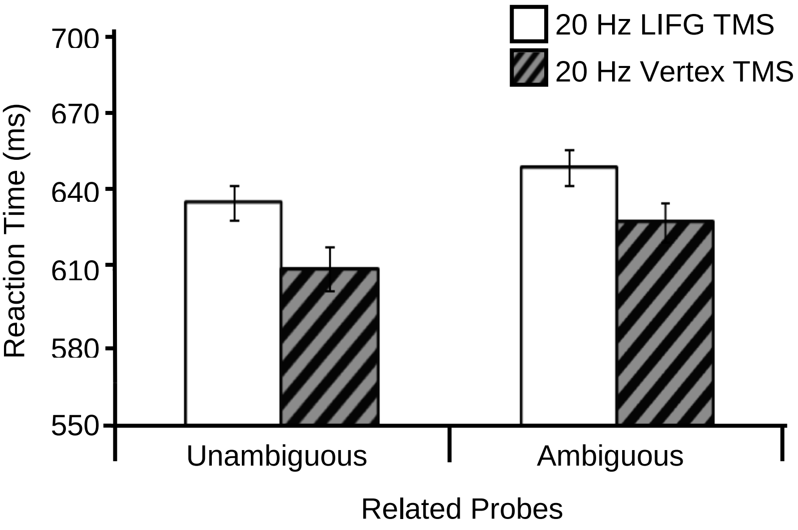

Figure 5-11: Mean reaction times for the related probes of ambiguous and unambiguous sentences, performed with 20 Hz LIFG TMS (white) and 20 Hz Vertex TMS (striped) stimulation. Error bars show standard error adjusted to remove between subject variance.

5.5.4 Discussion

The aim of this experiment was to re-investigate the effects that LIFG stimulation has on sentence comprehension using the modified set of TMS parameters (20 Hz for 150 ms). Experiment 9 showed that these parameters reduced the non-site-specific effects of stimulation that interacted with the ambiguity manipulation in the task found with 10 Hz stimulation (Experiment 8). As a result, any effects of LIFG stimulation with these modified parameters can be more confidently interpreted as resulting from disruptions to critical comprehension-related processes occurring within the LIFG, since they are not confounded with artefactual performance in the comparison site (Vertex) condition.

The results showed that participants were slower to decide that a probe was semantically related to the previously heard sentence, regardless of ambiguity, when stimulation was delivered to the LIFG than to Vertex. This LIFG effect is particularly convincing because performance in the Vertex control condition was comparable to that in the behavioural study in that ambiguity effects were found in both the accuracy and reaction time data. Although these ambiguity effects in the baseline Vertex data were only marginally significant, this is not unsurprising as the behavioural study (Chapter 2, Experiment 3) demonstrated that such effects are more robust with shorter probe delays. It was more critical that they were in the same direction as those of the behavioural study. Therefore, together, the control site data indicate that it is unlikely, in this experiment, that the LIFG effects are confounded by non-site-specific effects of stimulation but, instead, reflect genuine effects of disruption to LIFG functioning on sentence comprehension. Hence, these findings support the change in stimulation parameters from the first TMS experiment and confirm the Vertex condition as an appropriate control against which to compare the LIFG effects.

As found in the first TMS experiment, the disruptive effects of stimulation were not restricted to the ambiguous sentences. These findings converge on the conclusion that the LIFG serves a critical and routine role in comprehending the meaning of sentences. Again (see discussion of Experiment 5), this is highly compatible with sentence general theories of LIFG functioning, such as Hagoort's semantic unification account (Hagoort, 2005, 2013) whilst being more difficult to reconcile with sentence-specific theories, such as conflict resolution accounts (Novick et al., 2009; Novick et al., 2005). In addition, this pattern is, again, inconsistent with the results of the fMRI experiment presented in Chapter 4 and various other neuroimaging studies which show no significant LIFG response to unambiguous sentences (e.g., Crinion et al., 2003; Friederici et al., 2003; Rodd, Johnsrude, et al., 2012; Scott et al., 2000; Spitsyna et al., 2006). Such a finding highlights the mismatch that can be found between different neuroscientific techniques, which may be attributable to differences in their physiological basis (e.g., neuronal spikes vs. blood flow) or experimental factors (e.g., baseline) and emphasises the difficulty in interpreting null results (see Chapter 4 for further discussion). Thus, in contrast to conclusions drawn in many of these fMRI studies, the TMS LIFG effect provides positive evidence that this region is important for sentence comprehension in general.

The results, however, are inconclusive with regard to the LIFG's role in ambiguity resolution because there was no interaction between LIFG disruption and sentence ambiguity. That is, there was no evidence that LIFG stimulation differently affected the ambiguous sentences compared to the unambiguous sentences. The simplest interpretation of these results is the LIFG serves sentence-general processes that occur in both types of sentences and that these processes are not more critical for comprehending ambiguous than unambiguous sentences. For example, TMS was administered at the end of both sentences, which is when various researchers argue that important integrational processes occur, known as sentence wrap-up (Just &

Carpenter, 1980). The role of the LIFG in integrational processes is consistent with the kinds of operations proposed in Hagoort’s theory of semantic unification, although Hagoort argues that ambiguous sentences place greater demand on these processes (Hagoort, 2005, 2013). The alternative possibility is that stimulating the LIFG disrupted functionally different processes in the two sentences, which manifested in equivalent behavioural slow-downs because reaction times are general indexes of processing efficiency. For example, it is possible that semantic reinterpretation was disrupted in the ambiguous sentences while wrap-up processes were disrupted in the unambiguous sentences. The behavioural data presented in Chapter 2 support the notion that reinterpretation occurs when TMS is administered at the offset of these ambiguous sentences. Under this view, the lack of interaction must then arise because sentence wrap-up occurs later in the ambiguous sentences such that it was not yet occurring, and thus not disrupted, when TMS was administered. Otherwise, if both processes were affected in the ambiguous sentences, then an additive effect of stimulation would be expected.

Although this current data cannot test these suggestions, some preliminary evidence supporting the dual-process hypothesis comes from comparing these results with those of the first TMS experiment (Chapter 4, Experiment 5). In contrast to this current experiment, Experiment 5 showed a clear trend for stimulation to affect ambiguous-related probes more than unambiguous-related probes. Although the results of Experiment 5 are likely to have been influenced by artefactual task performance caused by the specific type of TMS parameters that was employed (Chapter 5, Experiment 8), it is possible that the LIFG effects were also different because stimulation was delivered for a longer period of time (300 ms versus 150 ms). For example, the interaction may have arisen in that experiment because stimulating a 300 ms time-window may have been long enough to disrupt both reinterpretation and sentence wrap-up processes in the ambiguous sentences. Unfortunately, the two sets of data cannot be directly compared due to the differences in baseline performance and, thus, prevent any strong conclusions from being drawn. However, they highlight TMS as a useful tool for testing such temporal hypotheses. For example, the current set of parameters (20 Hz, 4 pulses, 150 ms) could be administered at different time-points during sentence comprehension to examine whether, and when, the LIFG is critical for ambiguous and unambiguous sentence processing. Their results may then help guide behavioural experiments to investigate the precise processes occurring during these time points, in order to be able to attribute specific functions (e.g., reinterpretation versus sentence wrap-up) to the regions that TMS shows are involved at these times.

These temporal-based investigations that are afforded by TMS are also important because they may reveal transient involvements of a brain region in a particular process, to which other methods such as fMRI, may be less sensitive. For instance, while these TMS data showed evidence that the LIFG is necessary for unambiguous sentence processing, the fMRI study did not reveal a significant BOLD response to these sentences in this region (Chapter 3). While the fMRI null-result may be attributed to baseline effects (see above), it may also be possible that the LIFG is only transiently involved at certain points during comprehension, such as at the end of the sentence. Thus, the fMRI study may not have detected this response because the whole sentence was modelled in the analyses (Chapter 3).

The lack of interaction leaves open another more trivial interpretation. Responses may have been slower under LIFG stimulation because TMS to this site is more distracting than Vertex stimulation since it can produce some stimulation of facial nerves or muscles surrounding the LIFG site. This is unlikely to be a major contributing factor for a few reasons. First, stimulation was never presented at the time when the probe was presented and was only presented in the first 150 ms of the probe delay so cannot have any direct influence on the probe response. Second, the first TMS experiment (Chapter 4, Experiment 5) showed that LIFG stimulation did not all affect probe conditions (i.e., there was a probe condition x stimulation site interaction), suggesting that behavioural disruptions are not an inevitable outcome of LIFG stimulation. Furthermore, various other TMS studies that have controlled for the sensation of TMS on the LIFG have also found that disruption is specific to the task and to the area of the LIFG that is stimulated, which cannot be attributed to non-specific effects of distraction (Devlin et al., 2003; Gough et al., 2005). Nevertheless, future experiments need to include control conditions to directly examine the contribution of distraction to the LIFG effects.

In contrast to the first TMS experiment, the effects of LIFG stimulation were limited to reaction times. LIFG stimulation showed no evidence of affecting task accuracy, indicating that disrupting LIFG function did not prevent comprehension but affected the efficiency of this process. This is consistent with the pattern of result in Experiment 7 where 10 Hz stimulation was employed. As discussed there, this may have resulted from removing particularly difficult items, making reaction times the more informative and sensitive measure of LIFG disruption.

In summary, these results converge on the conclusion of the first TMS experiment, that the LIFG has a critical role in the semantic processing of both ambiguous and unambiguous sentences. These findings clearly indicate that a theory that focuses on ambiguity (or conflict) resolution is inadequate and provide support for a more general role of the LIFG in sentence comprehension, such as Hagoort's semantic unification account, where the LIFG serves to unify semantic information into larger representations of a sentence or discourse (Hagoort, 2005, 2013). However, the finding of a main effect of stimulation in the absence of an interaction raises several questions about the exact nature of these effects. While the current data cannot test alternative explanations of these results, the set of TMS experiments demonstrates a useful TMS paradigm that can be used to investigate these questions. For example, further insights into the exact role of this region may be gained by re-investigating the full set of probe conditions (unrelated and inappropriate probe conditions), as well as varying the time when stimulation is administered during the task. Importantly, this last experiment has demonstrated that, with carefully chosen and examined stimulation parameters, TMS can be successfully administered during spoken sentence comprehension without affecting behavioural performance, enabling the assessment of causal relationships between specific brain regions and sentence processing.

Chapter 6 General Discussion

6.1 Overview

The aim of this thesis was to increase our understanding of the cognitive and neural processes underlying the comprehension of semantically ambiguous words that are encountered during spoken sentences. More specifically, this thesis focused on how listeners reinterpret the meaning of these words when they encounter context late on in the sentence. The process of semantic reinterpretation has received little research attention, despite providing a window into various key aspects of language comprehension, including semantic flexibility, the influence of context on meaning comprehension and the involvement of high-level cognitive abilities in language processing such as selection, inhibition, and integration. Together, this thesis has revealed various cognitive and neural effects of semantic reinterpretation that extend understanding into ambiguity resolution as well as have implications for models of language comprehension more generally.

6.2 Cognitive Mechanisms Underlying Semantic Ambiguity Resolution

The first part of this thesis (Chapter 2) investigated the cognitive processes underlying semantic reinterpretation by employing a semantic-relatedness judgement paradigm that allowed the examination of two important aspects of reinterpretation that have not been and/or cannot be investigated with previous methods (i.e., eye-movement and non-linguistic dual-tasks). Specifically, the paradigm allowed the simultaneous investigation of 1) the effectiveness of reinterpretation and 2) the efficiency of this process, which, importantly, included being able to disentangle the relative time course of (re)activating the contextually appropriate meaning and suppressing the contextually inappropriate meanings.

The accuracy results showed that participants were highly effective at reinterpreting their understanding of ambiguous words in line with contextual information that was encountered at the end of the sentence. Importantly, they were highly accurate at both deciding that probes related to the contextually appropriate meaning of the ambiguous word (e.g., “speech”) were related to the sentence (e.g., “the woman had to make the toast with a very old *microphone*”) as well as in deciding that words which were associated with the inappropriate meaning (e.g., “bread”) were not related to the sentence. Together, these results demonstrate that semantic representations generated during comprehension can be flexibly and fully revised.

The reaction time data provided complementary findings about the efficiency of this process. First, the results showed that reinterpreting the meaning of a sentence carried a processing cost, as evinced by the finding that participants were slower to make a semantic judgement about the meaning of ambiguous sentences than unambiguous sentences when comprehension was probed 100 ms after the sentence. This behavioural cost concurs with the majority of previous research demonstrating that semantic reinterpretation is a time-consuming process. Second, the results revealed novel insights into the nature of these processing costs. Participants had difficulty in both accepting related probes (“speech”) as well as in rejecting inappropriate probes (“bread”) with this probe delay, indicating that integrating the contextually appropriate meaning as well as suppressing the contextually inappropriate meaning are both time-consuming. However, these RT costs were differently affected by the length of the probe delay, suggesting a computational dissociation between these two processes. Specifically, the results suggested that the suppression of inappropriate meanings is delayed relative to the integration of the contextually appropriate meanings.

Together, the results are particularly consistent with constraint-based theories of sentence processing, wherein comprehension is a dynamic process in which contextual (and other) information continually modulate comprehension but whose impact is dependent on how such information is weighted relative to others sources (MacDonald et al., 1994; MacDonald & Seidenberg, 2006). However, the way by which such information is weighted and how it affects different meanings is uncertain and in need for further investigation. The particular direction of the meaning dissociation that was found (i.e., delayed suppression of the inappropriate meaning) supports two potential alternatives. It is consistent with Gernsbacher and St John's (2002) model in which suppression is triggered by sentence-level representations rather than by the activation of the correct meaning of the ambiguous word per se. Alternatively, this dissociation may merely arise because the contextually inappropriate meaning is strongly weighted when the contextual information is encountered having been part of the initial interpretation of the sentence such that it takes more processing cycles to reduce its activation (MacDonald et al., 1994).

6.3 Neural Mechanisms Underlying Semantic Ambiguity Resolution

The second part of the thesis (Chapter 3, 4 and 5) examined the neural processes underlying the comprehension of the sentences investigated in Chapter 2. First, an fMRI study was conducted to investigate the network of brain regions that are engaged when listeners process these sentences that also examined the functional contributions of these ambiguity-related regions as well as the inter-subject variability surrounding these responses. The results demonstrated that these sentences produced greater activity in the mid-LIFG and left posterior inferior temporal cortex than unambiguous sentences, which broadly replicated the results of previous studies. The functional properties of these regions were further examined by investigating the response of these regions to various sentence conditions. First, there was no significant evidence that these regions responded to sentences that did not contain ambiguous words, suggesting that these specific regions of the frontal and temporal cortex may not be routinely involved in sentence processing. Second, both regions' responses showed evidence of being modulated by meaning dominance, such that responses were greater for sentences in which reinterpretation, rather than initial selection, was more demanding (i.e., biased > balanced). Since this dominance pattern is a fundamental prediction of ambiguity models, it constitutes key evidence that these regions play a role in semantic reinterpretation.

Furthermore, the inter-individual variability analysis revealed that both of these regions responded to ambiguity in nearly all participants, which further highlights them as fundamental regions associated with semantic reinterpretation. Interestingly, participants' responses were more anatomically consistent around the frontal group-level peak than the temporal group-peak, such that the majority of participants showed a peak that was located in pars triangularis. This finding, therefore, further suggested that the mid-LIFG may be particularly important for ambiguity resolution. Thus, the final part of this thesis (Chapter 4 and 5) examined the role of the LIFG further by using TMS to investigate whether this region has a causal role in semantic reinterpretation.

The first TMS experiment (Chapter 4) employed the behavioural paradigm presented in Chapter 2, as it was shown to be sensitive to the ambiguity of the preceding sentence. The findings showed preliminary evidence that the LIFG serves a necessary function for reinterpreting the meaning of sentences. Interestingly, it specifically showed that this region may support the integration of the contextually appropriate meaning since LIFG stimulation slowed participants decisions to ambiguous-related probes but not to the other ambiguous probes (i.e., ambiguous-inappropriate or ambiguous-unrelated), compared to when a control site was stimulated. However, two unexpected findings also came out of the data. First, the results showed that LIFG stimulation also affected unambiguous sentence comprehension, causing participants to be less accurate for unambiguous-related probes and slower to respond to unambiguous-unrelated probes. This appeared to contradict the results of the fMRI study in which the LIFG showed no significant response to unambiguous sentences (compared to a low-level auditory baseline). The second surprise was that the behavioural ambiguity effects found under control site (Vertex) stimulation were different than those in the previously-conducted behavioural experiment (Chapter 2, Experiment 2). This suggested that the LIFG effects may have been confounded by non-site-specific effects of TMS on task performance occurring in the control condition. As a result, the last part of the thesis conducted a set of TMS experiments to examine potential reasons behind these unexpected findings. Together, this final body of work revealed evidence that the mere act of administering TMS, regardless of head site, affected task performance in such a way that it facilitated responses to ambiguous-related probes more than unambiguous-related probes. This highlighted the need to change the TMS parameters to find a set that did not interact with ambiguity in order to be confident that any effects of LIFG stimulation would not be confounded by non-site-specific effects of stimulation. Increasing the rate of stimulation from 10 Hz to 20 Hz eliminated the specific effect that TMS had on ambiguous sentence processing (Chapter 5, Experiment 8), perhaps because this higher frequency changed the rhythmic sound of the TMS clicks and ensured that the last pulse did not coincide with, and thus cue, the occurrence of the probe word.

The final experiment (Chapter 5, Experiment 9) employed the improved set of TMS parameters and, like the first TMS experiment (Chapter 4) revealed a significant effect of LIFG stimulation on both ambiguous and unambiguous sentence comprehension. This supported the conclusions of the first experiment that the LIFG has a necessary role in the comprehension of both types of sentences and, thus, supports sentence-general theories of LIFG function such as the semantic unification account (Hagoort, 2005, 2013). It is, therefore, inconsistent with sentence-specific theories such as the conflict resolution account which argues that the LIFG supports sentence comprehension only when there is representational conflict (Novick et al., 2009; Novick et al., 2005). The “unambiguous LIFG effect” and, ensuing conclusion is different to that drawn from the fMRI results which showed no significant response to unambiguous sentences compared to a low-level auditory baseline. The TMS effect suggests that the lack of an effect in the fMRI study may have resulted from reduced sensitivity to responses to the unambiguous sentences. The fMRI null-result, for example, may have been concealed by activation in the baseline condition. Together the TMS and fMRI results, thus, highlight the difficulty in interpreting null-result and thus the importance of using converging methods for testing assumption of neural models of language and cognition in general. However, there was no evidence that the LIFG differently affected the ambiguous sentences, which leaves the results inconclusive regarding the precise role of the LIFG in ambiguity resolution. The main effect of stimulation may only indicate that this region is not specific for ambiguity resolution but cannot indicate whether stimulation affected functionally different processes in the ambiguous and unambiguous sentences such that, for example, reinterpretation processes were affected in the former. Clearly, future research is needed to investigate this hypothesis further. The TMS paradigm developed in this thesis is a useful tool to examine some of these outstanding questions, as it can be modified in various ways to examine whether, when and how the LIFG is involved in sentence comprehension more generally by, for example, manipulating the time when TMS is administered during sentence processing and examining different types of sentences and probe conditions. Importantly, the TMS experiments presented here highlight the need to pilot any subtle changes in the TMS parameters (timing, frequency) to ensure that they do not produce non-site-specific effects on task performance.

A methodological strength of this thesis is that the same stimuli were examined across experiments and, in addition, the same task was used in the behavioural and TMS studies. This methodological choice enhances the interpretation of the cognitive and neural effects. For example, this enables the behavioural study to provide support for the conclusions drawn in the fMRI study about the cognitive operations imputed to the ambiguity-elevated neural responses of the LIFG and posterior inferior temporal cortex. In addition, using the same behavioural paradigm in the TMS experiment as used in the behavioural study enabled the direct comparability of results, which helped highlight discrepancies that were due to methodological artefacts (e.g., the TMS artefact in Chapter 4).

6.4 Limitations and Future Directions

The results of this thesis, also, highlighted various unanswered questions for further research. Four important questions are considered here.

1) What is semantic reinterpretation?

The behavioural study (Chapter 2) showed that semantic reinterpretation incurs processing costs and extended previous research by demonstrating that these costs are associated with integrating the contextually appropriate meaning as well as suppressing inappropriate meanings. However, it is unknown what are the precise processes operating on these meanings. One of the reasons for this is that it is uncertain what “selection” entails. The assumption of current models of semantic ambiguity resolution is that when semantically ambiguous words are encountered prior to biasing contextual information one meaning is selected relatively quickly after encountering that word (Duffy et al., 2001; Seidenberg et al., 1982; Swinney, 1979; Twilley & Dixon, 2000). However, as discussed in Chapter 2, it is unclear what are the states of contextually appropriate meanings that are not initially selected and the states of contextually inappropriate meanings that are selected. If non-selected meanings are maintained at a lower level of activation (McRae et al., 1998), then reinterpretation would entail boosting their activation. If these meanings are suppressed/decay to baseline levels by the time that context is encountered (Duffy et al., 2001; Gernsbacher, 1990; Gernsbacher & St John, 2001; MacDonald et al., 1994), then reinterpretation would also entail “accessing” them. It is likely that these alternatives are dependent on the dominance of an ambiguous word’s meaning. For example, eye-movement research, in particular, has presented a large body of evidence showing that, if words have a more dominant meaning, then the subordinate meanings may be quickly disregarded, whereas words with more balanced meanings may be maintained for longer or to a higher level of activation (Duffy et al., 2001; Duffy et al., 1988; Gernsbacher & St John, 2001; Twilley & Dixon, 2000). Unfortunately, it was not possible to test effects of dominance in the behavioural study (Chapter 2) as there were too few items per dominance category when probe condition was factored in. However, dominance effects were able to be tested in the fMRI data as no probe task was included so the sentence trials did not need to be divided across probe conditions. These results showed that meaning dominance modulated neural responses, suggesting that reinterpretation processes may be functionally different depending on the frequency of an ambiguous word’s meaning.

There is more agreement about the state of initially-selected meanings: these meanings are integrated into sentence-level representations (Duffy et al., 2001; Duffy et al., 1988; Gernsbacher & St John, 2001; MacDonald et al., 1994; Seidenberg et al., 1982). Thus, reinterpretation must involve suppressing this high-level sentence representation but the unresolved question is whether this is independent of suppressing the inappropriate meaning? The behavioural study

(Chapter 2) demonstrated that the inappropriate meaning still causes interference several hundred milliseconds after inconsistent context is encountered. But is this because it takes time to suppress the inappropriate meaning itself or to suppress the sentence-level representation that encompasses that meaning? In addition, the latter results also raise the question of why suppression of the inappropriate meaning is delayed. Is this an inevitable consequence of their strong level of activation (Gernsbacher & St John, 2001; MacDonald et al., 1994) or does this reflect a more deliberate process, where incorrect meanings are maintained (or slowly reduced) in case revision is needed later on (Faust & Chiarello, 1998)?

2) How does context trigger reinterpretation?

Another important question is how context instigates reinterpretation: how direct is its influence? Reinterpretation may be initiated because contextual information activates semantic representations that directly conflict with the initially selected meaning of the ambiguous word and/or it may be initiated more indirectly by way of activating the alternative meaning of the ambiguous word which competes with the inappropriate meaning. The possibility of different mechanisms is apparent in various models of ambiguity resolution. For example, an indirect process is suggested by Duffy et al.’s (1988) reordered access model, which assumes that context only has a facilitatory role and, therefore, cannot suppress the inappropriate meanings directly. In contrast, constraint-based connectionist models assume that multiple constraints, including contextual information, can directly inhibit and facilitate alternative meanings (MacDonald et al., 1994). One way that these hypotheses can be examined is by assessing the time-course of their activation with more sensitive measures of the activation levels of the appropriate and inappropriate meanings. The behavioural study employed an explicit semantic judgement task in which participants needed to judge the semantic relationship between the sentence and probe word. This complex decision may have masked subtle differences in activation levels. That is, although the results showed that the inappropriate meanings are suppressed more slowly than the appropriate meanings are integrated, they are inconclusive as to whether this is because the former has a slower time-course, where the decrements are too small to be detected by the explicit-semantic task, or because this process was initiated later (i.e., as by the appropriate meaning). More sensitive measures may be non-explicit semantic tasks, such as lexical decision (Klepousniotou, 2002; Seidenberg et al., 1982; Swinney, 1979) or event-related potentials that are measured whilst participant passively read the different probe words (Swaab et al., 2003) as they reduce the decision components.

3) How should future research take into account individual differences in meaning preferences?

This thesis examined semantic reinterpretation by employing sentences in which ambiguous words were disambiguated to their subordinate meaning at the end of the sentence. The decision about which meaning was subordinate was initially based on meaning dominance scores conducted on a similar set of stimuli (Rodd, Johnsrude, et al., 2010) and were then checked for the specific set of sentences used in this thesis via word association and sentence-completion tasks (see Chapter 2, Experiment 1 method section). However, a dominance score is the average preference of a separate group of participants who do not take part in the main experiment and, thus does not take into individual differences in meaning preferences. Nevertheless, it is currently the standard way to determine the relative frequencies of a word’s meanings in semantic ambiguity research (Colbert-Getz & Cook, 2013; Leinenger & Rayner, 2013; Seidenberg et al., 1982; Sereno et al., 2006; Swinney, 1979). The rationale behind using average preference scores is two-fold: 1) they are an efficient index of meaning frequency as individual participant’s preferences do not need to be measured and 2) they have been validated in many studies, in that behavioural effects of meaning frequency that are based on these scores are found (e.g., Duffy et al., 2001; Simpson, 1994; Twilley & Dixon, 2000). The potential problem with averages is that not all individuals have the same preferences, such that for some people the “subordinate” meaning may actually be their preferred meaning. As a result, some subtle effects of dominance may be potentially masked by trials in which the subordinate meaning is not employed for certain participants. This may have contributed to the failure to find dominance effects in the behavioural study and at the whole brain level in the fMRI study. Therefore, it would be useful for future research to either acquire meaning preferences for each individual basis prior to the main experiment or explicitly control individual’s preference, for example, using a priming task before the main task.

4) What are the specific functions of regions in semantic ambiguity resolution?

The results also raise questions regarding the neural processes underlying ambiguity resolution. The fMRI results demonstrated that the LIFG and posterior temporal cortex are both engaged by semantic reinterpretation but the current results cannot indicate the functional relationship between these regions. How do they interact during this process? Does information flow from the temporal cortex to the inferior frontal cortex or vice versa, or are they performing independent functions? These questions need to be investigated with methods that have high temporal resolution, such as MEG, or with causal modelling of BOLD responses, such as dynamic causal modelling (DCM). The TMS paradigm developed in this thesis is another potentially useful tool to investigate the temporal relationship between these different regions, since TMS can be administered at different time-windows during sentence comprehension. The final TMS experiment (Chapter 5, Experiment 10) showed that TMS can produce significant effects of LIFG stimulation on sentence processing, suggesting that this region may have a necessary role in this process. However, it was also clear that the some further modifications to this paradigm are needed before strong conclusions can be drawn from its results. In particular, control tasks and conditions need to be included in the TMS paradigm that do not require processes associated with the region under investigation in order to be able to assess whether the stimulation effects reflect non-specific effects that are not related to disturbing processing in that cortical region (e.g., distraction by facial nerve stimulation). In addition, the TMS results did not show conclusive evidence that this region serves a causal role in ambiguity resolution, since stimulating the LIFG did not differently affect both ambiguous and unambiguous sentence processing. However, it did provide a methodological foundation for future research to use this paradigm to tease apart the role of this region (and other regions) in ambiguity resolution and sentence comprehension more generally. For example, future experiments could use this paradigm to assess whether specific regions have different temporal involvements in the processing of these two types of sentences.

Furthermore, the results also cannot identify the specific functions of these regions during semantic reinterpretation. The neural findings show activation in regions that have been attributed to various different functions, including conflict resolution, semantic unification, visual imagery, sound-to-meaning mapping, but currently there has been no direct investigation about the relationship between these functions and ambiguity resolution. Such a knowledge gap highlights the need for future psycholinguistic work to go beyond viewing semantic reinterpretation as an interaction between integration and suppression of meanings and investigate these other processes which, in turn, may help specify the roles that ambiguity-elevated regions are serving in ambiguity resolution. Thus, this further highlights the need for psycholinguistic and neural research to continue guiding each other in order to achieve a comprehensive understanding of ambiguity resolution and language comprehension in general.

References

Acheson, D. J., & Hagoort, P. (2013). Stimulating the Brain's Language Network: Syntactic Ambiguity Resolution after TMS to the Inferior Frontal Gyrus and Middle Temporal Gyrus. *Journal of Cognitive Neuroscience, 25*(10), 1664-1677.

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *Neuroimage, 49*(1), 1124-1132.

Antal, A., Terney, D., Poreisz, C., & Paulus, W. (2007). Towards unravelling task-related modulations of neuroplastic changes induced in the human motor cortex. *Eur J Neurosci, 26*(9), 2687-2691.

Ashbridge, E., Walsh, V., & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia, 35*(8), 1121-1131.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM).* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron, 47*(6), 907-918.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia, 45*(13), 2883-2901.

Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology, 20*(5), 529.

Balota, D. A., Black, S. R., & Cheney, M. (1992). Automatic and attentional priming in young and older adults: Reevaluation of the two-process model. *Journal of Experimental Psychology: Human Perception and Performance, 18*(2), 485.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion-symptom mapping. *Nat Neurosci, 6*(5), 448-450.

Baumgaertner, A., Weiller, C., & Büchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *Neuroimage, 16*(3), 736-745.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience, 15*(7), 991-1001.

Beck, D. M., Muggleton, N., Walsh, V., & Lavie, N. (2006). Right parietal cortex plays a critical role in change blindness. *Cereb Cortex, 16*(5), 712-717.

Bedny, M., Hulbert, J. C., & Thompson-Schill, S. L. (2007). Understanding words in context: The role of Broca's area in word comprehension. *Brain Research, 1146*(1), 101-114.

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

Bekinschtein, T. A., Davis, M. H., Rodd, J. M., & Owen, A. M. (2011). Why Clowns Taste Funny: The Relationship between Humor and Semantic Ambiguity. *Journal of Neuroscience, 31*(26), 9665-9671.

Binder, J., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767-2796.

Binder, J., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state. A functional MRI study. *J Cogn Neurosci, 11*(1), 80-95.

Binder, J., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10*(5), 512-528.

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.

Binder, K. S., & Morris, R. K. (2011). An Eye-Movement Analysis of Ambiguity Resolution: Beyond Meaning Access. *Discourse Processes, 48*(5), 305-330.

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.

Boatman, D., Gordon, B., Hart, J., Selnes, O., Miglioretti, D., & Lenz, F. (2000). Transcortical sensory aphasia: revisited and revised. *Brain, 123*(8), 1634-1642.

Brasil-Neto, J. P., Cohen, L. G., Panizza, M., Nilsson, J., Roth, B. J., & Hallett, M. (1992). Optimal focal transcranial magnetic activation of the human motor cortex: effects of coil orientation, shape of the induced current pulse, and stimulus intensity. *J Clin Neurophysiol, 9*(1), 132-136.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage, 16*, S497.

Broca, P. (1861). Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau. *Bull Soc Anthropol, 2*, 235-238.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: a combined TMS and behavioral study. *Brain Res Cogn Brain Res, 24*(3), 355-363.

Burgess, C., Tanenhaus, M. K., & Seidenberg, M. S. (1989). Context and lexical access: Implications of nonword interference for lexical ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 620.

Caplan, D., Alpert, N., & Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience, 10*(4), 541-552.

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain, 119*(3), 933-949.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of Algorithmic and Heuristic Processes in Language Comprehension - Evidence from Aphasia. *Brain and Language, 3*(4), 572-582.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*(8), 1839-1849.

Chan, D., Fox, N. C., Scahill, R. I., Crum, W. R., Whitwell, J. L., Leschziner, G., et al. (2001). Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Annals of neurology, 49*(4), 433-442.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex, 12*(5), 545-551.

Chen, L., & Boland, J. E. (2008). Dominance and context effects on activation of alternative homophone meanings. *Memory & cognition, 36*(7), 1306-1323.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology, 42*(4), 368-407.

Civardi, C., Cantello, R., Asselman, P., & Rothwell, J. C. (2001). Transcranial magnetic stimulation can be used to test connections to primary motor areas from frontal and medial cortex in humans. *Neuroimage, 14*(6), 1444-1453.

Cohen, L., Jobert, A., Le Bihan, D., & Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *NeuroImage, 23*(4), 1256-1270.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain, 125*(5), 1054-1069.

Colbert-Getz, J., & Cook, A. E. (2013). Revisiting effects of contextual strength on the subordinate bias effect: Evidence from eye movements. *Memory & cognition*, 1-13.

Coleman, M. R., Davis, M. H., Rodd, J. M., Robson, T., Ali, A., Owen, A. M., et al. (2009). Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain, 132*, 2541-2552.

Coleman, M. R., Rodd, J. M., Davis, M. H., Johnsrude, I. S., Menon, D. K., Pickard, J. D., et al. (2007). Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. *Brain, 130*, 2494-2507.

Corbett, F., Jefferies, E., Ehsan, S., & Lambon Ralph, M. A. (2009). Different impairments of semantic cognition in semantic dementia and semantic aphasia: evidence from the non-verbal domain. *Brain, 132*(9), 2593-2608.

Corbett, F., Jefferies, E., & Lambon Ralph, M. A. (2009). Exploring multimodal semantic control impairments in semantic aphasia: Evidence from naturalistic object use. *Neuropsychologia, 47*(13), 2721-2731.

Corbett, F., Jefferies, E., & Lambon Ralph, M. A. (2011). Deregulated semantic cognition follows prefrontal and temporo-parietal damage: Evidence from the impact of task constraint on nonverbal object use. *Journal of Cognitive Neuroscience, 23*(5), 1125-1135.

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. S. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain, 126*(5), 1193-1201.

D'Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., et al. (1997). A functional MRI study of mental image generation. *Neuropsychologia, 35*(5), 725-730.

D’Esposito, M., Postle, B. R., Jonides, J., & Smith, E. E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proceedings of the National Academy of Sciences, 96*(13), 7514-7519.

Damasio, A. R. (1991). Signs of aphasia. *Acquired aphasia*, 27-43.

Damasio, H., Grabowski, T. J., Tranel, D., & Hichwa, R. D. (1996). A neural basis for lexical retrieval. *Nature*.

Damasio, H., Grabowski, T. J., Tranel, D., Ponto, L. L. B., Hichwa, R. D., & Damasio, A. R. (2001). Neural correlates of naming actions and of naming spatial relations. *Neuroimage, 13*(6), 1053-1064.

Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron, 24*(2), 427-432.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., et al. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences of the United States of America, 104*(41), 16032-16037.

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

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in cognitive sciences, 15*(6), 254-262.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience, 18*(1), 193-222.

Devlin, J. T., Jamison, H. L., Gonnerman, L. M., & Matthews, P. M. (2006). The role of the posterior fusiform gyrus in reading. *Journal of cognitive neuroscience, 18*(6), 911-922.

Devlin, J. T., Matthews, P. M., & Rushworth, M. F. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J Cogn Neurosci, 15*(1), 71-84.

Devlin, J. T., Moore, C. J., Mummery, C. J., Gorno-Tempini, M. L., Phillips, J. A., Noppeney, U., et al. (2002). Anatomic constraints on cognitive theories of category specificity. *Neuroimage, 15*(3), 675-685.

Devlin, J. T., Rushworth, M. F., & Matthews, P. M. (2005). Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia, 43*(1), 69-74.

Devlin, J. T., & Watkins, K. E. (2007). Stimulating language: insights from TMS. *Brain, 130*(3), 610-622.

Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., & Cabanis, E. A. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain, 130*(Pt 5), 1432-1441.

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*(1-2), 145-177.

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*. 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.

Duncan, K. J., Pattamadilok, C., & Devlin, J. T. (2010). Investigating Occipito-temporal Contributions to Reading with TMS. *Journal of Cognitive Neuroscience, 22*(4), 739-750.

Duncan, K. J., Pattamadilok, C., Knierim, I., & Devlin, J. T. (2009). Consistency and variability in functional localisers. *Neuroimage, 46*(4), 1018-1026.

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation. *Brain Res Rev, 62*(2), 233-244.

Fallon, M., Peelle, J. E., & Wingfield, A. (2006). Spoken sentence processing in young and older adults modulated by task demands: evidence from self-paced listening. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 61*(1), P10-P17.

Faust, M. E., & Chiarello, C. (1998). Sentence context and lexical ambiguity resolution by the two hemispheres. *Neuropsychologia, 36*(9), 827-835.

Faust, M. E., & Gernsbacher, M. A. (1996). Cerebral mechanisms for suppression of inappropriate information during sentence comprehension. *Brain Lang, 53*(2), 234-259.

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.

Foss, D. J. (1970). Some Effects of Ambiguity Upon Sentence Comprehension. *Journal of Verbal Learning and Verbal Behavior, 9*(6), 699-&.

Franzmeier, I., Hutton, S. B., & Ferstl, E. C. (2012). The role of the temporal lobe in contextual sentence integration: A single-pulse transcranial magnetic stimulation study. *Cogn Neurosci, 3*(1), 1-7.

Fridriksson, J., Bonilha, L., & Rorden, C. (2007). Severe Broca's aphasia without Broca's area damage. *Behav Neurol, 18*(4), 237-238.

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in cognitive sciences, 16*(5), 262-268.

Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex, 13*(2), 170-177.

Friston, K., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., & Frackowiak, R. S. (1995). Spatial registration and normalization of images. *Human brain mapping, 3*(3), 165-189.

Fuggetta, G., Pavone, E. F., Walsh, V., Kiss, M., & Eimer, M. (2006). Cortico-cortical interactions in spatial attention: a combined ERP/TMS study. *Journal of neurophysiology, 95*(5), 3277-3280.

Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the national Academy of Sciences, 95*(3), 906-913.

Ganaden, R. E., Mullin, C. R., & Steeves, J. K. (2013). Transcranial magnetic stimulation to the transverse occipital sulcus affects scene but not object processing.

Garrard, P., & Hodges, J. (2000). Semantic dementia: clinical, radiological and pathological perspectives. *Journal of neurology, 247*(6), 409-422.

Gennari, S. P., MacDonald, M. C., Postle, B. R., & Seidenberg, M. S. (2007). Context-dependent interpretation of words: Evidence for interactive neural processes. *Neuroimage, 35*(3), 1278-1286.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, N.J.: L. Erlbaum.

Gernsbacher, M. A., & Faust, M. E. (1991). The Mechanism of Suppression: A Component of General Comprehension Skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 245-262.

Gernsbacher, M. A., & Robertson, R. R. W. (1995). Reading skill and suppression revisited. *Psychological Science*, 165-169.

Gernsbacher, M. A., & St John, M. F. (2001). Modeling suppression in lexical access. *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity*, 47-65.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating Differences in General Comprehension Skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 430-445.

Geschwind, N. (1970). The organization of language and the brain. *Science, 170*(961), 940-944.

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.

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *The Journal of Neuroscience, 26*(24), 6523-6532.

Gold, B. T., & Buckner, R. L. (2002). Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron, 35*(4), 803-812.

Goodglass, H., Blumstein, S. E., Gleason, J. B., Hyde, M. R., Green, E., & Statlender, S. (1979). The effect of syntactic encoding on sentence comprehension in aphasia. *Brain and Language, 7*(2), 201-209.

Gough, P. M., Nobre, A. C., & Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *Journal of Neuroscience, 25*(35), 8010-8016.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci., 27*, 649-677.

Grindrod, C. M. (2012). Effects of left and right hemisphere damage on sensitivity to global context during lexical ambiguity resolution. *Aphasiology, 26*(7), 933-952.

Grindrod, C. M., & Baum, S. R. (2003). Sensitivity to local sentence context information in lexical ambiguity resolution: Evidence from left- and right-hemisphere-damaged individuals. *Brain and Language, 85*(3), 503-523.

Grodzinsky, Y. (1986). Language deficits and the theory of syntax. *Brain and language, 27*(1), 135-159.

Grodzinsky, Y., Piñango, M. M., Zurif, E., & Drai, D. (1999). The critical role of group studies in neuropsychology: Comprehension regularities in Broca's aphasia. *Brain and Language, 67*(2), 134-147.

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

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*(6), 883-899.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9*(9), 416-423.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Front Psychol, 4*, 416.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. *The cognitive neurosciences, 4*, 819-836.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438-441.

Hamilton, A. C., & Martin, R. C. (2005). Dissociations among tasks involving inhibition: A single-case study. *Cognitive, Affective, & Behavioral Neuroscience, 5*(1), 1-13.

Harpaz, Y., Levkovitz, Y., & Lavidor, M. (2009). Lexical ambiguity resolution in Wernicke's area and its right homologue. *Cortex, 45*(9), 1097-1103.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1), 67-99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Hilgetag, C. C., Theoret, H., & Pascual-Leone, A. (2001). Enhanced visual spatial attention ipsilateral to rTMS-induced 'virtual lesions' of human parietal cortex. *Nat Neurosci, 4*(9), 953-957.

Humphreys, G. F., & Gennari, S. P. (2014). Competitive mechanisms in sentence processing: Common and distinct production and reading comprehension networks linked to the prefrontal cortex. *NeuroImage, 84*, 354-366.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport, 12*(8), 1749-1752.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144.

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron, 28*(3), 979-990.

Jahanshahi, M., & Rothwell, J. (2000). Transcranial magnetic stimulation studies of cognition: an emerging field. *Exp Brain Res, 131*(1), 1-9.

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience, 21*(12), 2434-2444.

Jefferies, E. (2013). The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS. *Cortex, 49*(3), 611-625.

Jefferies, E., Baker, S. S., Doran, M., & Lambon Ralph, M. A. (2007). Refractory effects in stroke aphasia: A consequence of poor semantic control. *Neuropsychologia, 45*(5), 1065-1079.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain, 129*(8), 2132-2147.

Jefferies, E., Patterson, K., & Lambon Ralph, M. A. (2008). Deficits of knowledge versus executive control in semantic cognition: Insights from cued naming. *Neuropsychologia, 46*(2), 649-658.

Jonides, J., & Nee, D. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience, 139*(1), 181-193.

Juan, C. H., & Walsh, V. (2003). Feedback to V1: a reverse hierarchy in vision. *Experimental Brain Research, 150*(2), 259-263.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review, 87*, 329-354.

Kable, J. W., Lease-Spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *Journal of Cognitive Neuroscience, 14*(5), 795-805.

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.

Kan, I. P., Barsalou, L. W., Olseth Solomon, K., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology, 20*(3-6), 525-540.

Kan, I. P., & Thompson-Schill, S. L. (2004a). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience, 4*(1), 43-57.

Kan, I. P., & Thompson-Schill, S. L. (2004b). Selection from perceptual and conceptual representations. *Cognitive, Affective, & Behavioral Neuroscience, 4*(4), 466-482.

Kassuba, T., Klinge, C., Hölig, C., Menz, M. M., Ptito, M., Röder, B., et al. (2011). The left fusiform gyrus hosts trisensory representations of manipulable objects. *NeuroImage, 56*(3), 1566-1577.

Katzev, M., Tüscher, O., Hennig, J., Weiller, C., & Kaller, C. P. (2013). Revisiting the Functional Specialization of Left Inferior Frontal Gyrus in Phonological and Semantic Fluency: The Crucial Role of Task Demands and Individual Ability. *The Journal of Neuroscience, 33*(18), 7837-7845.

Kertesz, A., Sheppard, A., & MacKenzie, R. (1982). Localization in transcortical sensory aphasia. *Archives of Neurology, 39*(8), 475.

Kherif, F., Josse, G., & Price, C. J. (2011). Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cerebral Cortex, 21*(1), 103-114.

Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Reading anomalous sentences: an event-related fMRI study of semantic processing. *Neuroimage, 17*(2), 842-850.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review, 85*(5), 363.

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

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., et al. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science, 284*(5411), 167-170.

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage, 21*(3), 946-953.

Kuperberg, G. R., Sitnikova, T., & Lakshmanan, B. M. (2008). Neuroanatomical distinctions within the semantic system during sentence comprehension:

evidence from functional magnetic resonance imaging. *Neuroimage, 40*(1), 367-388.

Leinenger, M., & Rayner, K. (2013). Eye movements while reading biased homographs: Effects of prior encounter and biasing context on reducing the subordinate bias effect. *Journal of Cognitive Psychology, 25*(6), 665-681.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869-878.

Logothetis, N. K., Augath, M., Murayama, Y., Rauch, A., Sultan, F., Goense, J., et al. (2010). The effects of electrical microstimulation on cortical signal propagation. *Nat Neurosci, 13*(10), 1283-1291.

Lüders, H., Lesser, R., Hahn, J., Dinner, D., Morris, H., Wyllie, E., et al. (1991). Basal temporal language area. *Brain, 114*(2), 743-754.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution [corrected]. *Psychol Rev, 101*(4), 676-703.

MacDonald, M. C., & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics, 2nd Edition* (pp. 581-611). London: Elsevier Inc.

Mackay, D. G. (1966). To End Ambiguous Sentences. *Perception & Psychophysics, 1*(12), 426-436.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris, 102*(1), 59-70.

Mars, R. B., Klein, M. C., Neubert, F.-X., Olivier, E., Buch, E. R., Boorman, E. D., et al. (2009). Short-latency influence of medial frontal cortex on primary motor cortex during action selection under conflict. *The Journal of Neuroscience, 29*(21), 6926-6931.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol., 58*, 25-45.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current opinion in neurobiology, 11*(2), 194-201.

Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research, 1146*(1), 115-127.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*(3), 283-312.

Mellet, E., Tzourio, N., Denis, M., & Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *Neuroreport, 9*(5), 803-808.

Meltzer, J. A., Wagage, S., Ryder, J., Solomon, B., & Braun, A. R. (2013). Adaptive significance of right hemisphere activation in aphasic language comprehension. *Neuropsychologia*.

Milberg, W., Blumstein, S. E., & Dworetzky, B. (1987). Processing of lexical ambiguities in aphasia. *Brain Lang, 31*(1), 138-150.

Milham, M. P., Banich, M. T., & Barad, V. (2003). Competition for priority in processing increases prefrontal cortex's involvement in top-down control: an event-related fMRI study of the Stroop task. *Cognitive brain research, 17*(2), 212-222.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience, 24*(1), 167-202.

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working Memory Constraints on the Resolution of Lexical Ambiguity: Maintaining Multiple Interpretations in Neutral Contexts. *Journal of Memory and Language, 33*(2), 175-202.

Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia: pathologic and clinical. *Neurology, 28*(4), 311-324.

Moliadze, V., Zhao, Y., Eysel, U., & Funke, K. (2003). Effect of transcranial magnetic stimulation on single-unit activity in the cat primary visual cortex. *J Physiol, 553*(Pt 2), 665-679.

Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain, 122*(5), 943-962.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Annals of neurology, 47*(1), 36-45.

Mummery, C. J., Patterson, K., Wise, R. J. S., Vandenbergh, R., Price, C. J., & Hodges, J. R. (1999). Disrupted temporal lobe connections in semantic dementia. *Brain, 122*(1), 61-73.

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.

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.

Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., & Passingham, R. (2004). The Inferior Frontal Gyrus and Phonological Processing: An Investigation using rTMS. *Journal of Cognitive Neuroscience, 16*(2), 289-300.

Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature, 372*(6503), 260-263.

Noonan, K. A., Jefferies, E., Corbett, F., & Lambon Ralph, M. A. (2010). Elucidating the nature of deregulated semantic cognition in semantic aphasia: Evidence for the roles of prefrontal and temporo-parietal cortices. *Journal of Cognitive Neuroscience, 22*(7), 1597-1613.

Noonan, K. A., Jefferies, E., Visser, M., & Lambon Ralph, M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *Journal of cognitive neuroscience, 25*(11), 1824-1850.

Noppeney, U., Phillips, J., & Price, C. J. (2004). The neural areas that control the retrieval and selection of semantics. *Neuropsychologia, 42*(9), 1269-1280.

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.

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.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of cognitive neuroscience, 12*(6), 1013-1023.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex, 20*(3), 633-640.

Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I. P., Joensuu, R., Autti, T., et al. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage, 25*(2), 333-338.

Oliveri, M., Romero, L., & Papagno, C. (2004). Left but not right temporal involvement in opaque idiom comprehension: a repetitive transcranial magnetic stimulation study. *J Cogn Neurosci, 16*(5), 848-855.

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.

Parks, R., Ray, J., & Bland, S. (Eds.). (1998) Wordsmyth English dictionary-thesaurus. University of Chicago.

Pascual-Leone, A., Gates, J. R., & Dhuna, A. (1991). Induction of speech arrest and counting errors with rapid-rate transcranial magnetic stimulation. *Neurology, 41*(5), 697-702.

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering Misinterpretations in Garden-Path Sentences: Evidence From a

Paraphrasing Task. *Journal of Experimental Psychology: Learning Memory and Cognition, 35*(1), 280-285.

Paus, T., Jech, R., Thompson, C. J., Comeau, R., Peters, T., & Evans, A. C. (1997). Transcranial magnetic stimulation during positron emission tomography: a new method for studying connectivity of the human cerebral cortex. *The Journal of neuroscience, 17*(9), 3178-3184.

Penfield, W., & Roberts, L. (1959). Speech and brain mechanisms.

Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., et al. (1999). The neural correlates of verb and noun processing A PET study. *Brain, 122*(12), 2337-2344.

Persson, J., Sylvester, C.-Y. C., Nelson, J. K., Welsh, K. M., Jonides, J., & Reuter-Lorenz, P. A. (2004). Selection requirements during verb generation: differential recruitment in older and younger adults. *Neuroimage, 23*(4), 1382-1390.

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology, 19*(4), 319-324.

Pitcher, D., Goldhaber, T., Duchaine, B., Walsh, V., & Kanwisher, N. (2012). Two critical and functionally distinct stages of face and body perception. *J Neurosci, 32*(45), 15877-15885.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Exp Brain Res, 209*(4), 481-493.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology, 17*(18), 1568-1573.

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia, 48*(5), 1336-1342.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage, 10*(1), 15-35.

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.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310*(5756), 1963-1966.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage, 62*(2), 816-847.

Price, C. J., & Crinion, J. (2005). The latest on functional imaging studies of aphasic stroke. *Curr Opin Neurol, 18*(4), 429-434.

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *Neuroimage, 19*(3), 473-481.

Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences, 15*(6), 246-253.

Price, C. J., Moore, C. J., Humphreys, G. W., Frackowiak, R. S. J., & Friston, K. J. (1996). The neural regions sustaining object recognition and naming. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 263*(1376), 1501-1507.

Price, C. J., Noppeney, U., Phillips, J., & Devlin, J. T. (2003). How is the fusiform gyrus related to category-specificity? *Cognitive Neuropsychology, 20*(3-6), 561-574.

Price, C. J., Winterburn, D., Giraud, A. L., Moore, C. J., & Noppeney, U. (2003). Cortical localisation of the visual and auditory word form areas: a reconsideration of the evidence. *Brain and language, 86*(2), 272-286.

Price, C. J., Wise, R. J., & Frackowiak, R. S. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cereb Cortex, 6*(1), 62-70.

Price, C. J., Wise, R. J., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., et al. (1996). Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain, 119 (Pt 3)*, 919-931.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience, 12*(6), 718-724.

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.

Rayner, K., & Frazier, L. (1989). Selection Mechanisms in Reading Lexically Ambiguous Words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 779-790.

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A, 53*(4), 1061-1080.

Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of Prior Encounter and Global Discourse Bias on the Processing of Lexically Ambiguous Words: Evidence From Eye Fixations. *Journal of Memory and Language, 33*(4), 527-544.

Robinson, G., Blair, J., & Cipolotti, L. (1998). Dynamic aphasia: an inability to select between competing verbal responses? *Brain, 121*(1), 77-89.

Robinson, G., Shallice, T., & Cipolotti, L. (2005). A failure of high level verbal response selection in progressive dynamic aphasia. *Cognitive Neuropsychology, 22*(6), 661-694.

Robson, H., Sage, K., & Lambon Ralph, M. A. (2012). Wernicke's aphasia reflects a combination of acoustic-phonological and semantic control deficits: A case-series comparison of Wernicke's aphasia, semantic dementia and semantic aphasia. *Neuropsychologia, 50*(2), 266-275.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., et al. (2012). Learning new meanings for old words: effects of semantic relatedness. *Memory & Cognition, 40*(7), 1095-1108.

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.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89-104.

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (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.

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

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.

Rodd, J. M., Lopez Cutrin, B., 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.

Rogalsky, C., & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex, 19*(4), 786-796.

Rogalsky, C., Matchin, W., & Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: an fMRI study. *Frontiers in human neuroscience, 2*.

Rushworth, M. F., Paus, T., & Sipila, P. K. (2001). Attention systems and the organization of the human parietal cortex. *J Neurosci, 21*(14), 5262-5271.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M. S., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *Neuroimage, 27*(1), 188-200.

Sadeh, B., Pitcher, D., Brandman, T., Eisen, A., Thaler, A., & Yovel, G. (2011). Stimulation of category-selective brain areas modulates ERP to their preferred categories. *Current Biology, 21*(22), 1894-1899.

Sakai, K. L., Noguchi, Y., Takeuchi, T., & Watanabe, E. (2002). Selective priming of syntactic processing by event-related transcranial magnetic stimulation of Broca's area. *Neuron, 35*(6), 1177-1182.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences, 105*(46), 18035-18040.

Schofield, T. M., Penny, W. D., Stephan, K. E., Crinion, J., Thompson, A. J., Price, C. J., et al. (2012). Changes in auditory feedback connections determine the severity of speech processing deficits after stroke. *The Journal of Neuroscience, 32*(12), 4260-4270.

Schwartz, M. F., Saffran, E. M., & Marin, O. S. M. (1980). The word order problem in agrammatism. I. Comprehension. *Brain and Language, 10*(2), 249-262.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*(12), 2400-2406.

Seghier, M. L., & Price, C. J. (2011). Explaining left lateralization for words in the ventral occipitotemporal cortex. *The Journal of Neuroscience, 31*(41), 14745-14753.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*(4), 489-537.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre-and postlexical loci of contextual effects on word recognition. *Memory & Cognition, 12*(4), 315-328.

Sereno, S. C. (1995). Resolution of lexical ambiguity: evidence from an eye movement priming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 582.

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.

Sharp, D. J., Awad, M., Warren, J. E., Wise, R. J., Vigliocco, G., & Scott, S. K. (2010). The neural response to changing semantic and perceptual complexity during language processing. *Human brain mapping, 31*(3), 365-377.

Sheridan, H., & Reingold, E. M. (2012). The time course of contextual influences during lexical ambiguity resolution: Evidence from distributional analyses of fixation durations. *Memory & cognition, 40*(7), 1122-1131.

Sheridan, H., Reingold, E. M., & Daneman, M. (2009). Using puns to study contextual influences on lexical ambiguity resolution: Evidence from eye movements. *Psychonomic bulletin & review, 16*(5), 875-881.

Siebner, H. R., Hartwigsen, G., Kassuba, T., & Rothwell, J. C. (2009). How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex, 45*(9), 1035-1042.

Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior, 20*(1), 120-136.

Simpson, G. B. (1994). Context and the processing of ambiguous words. *Handbook of psycholinguistics, 22*, 359-374.

Simpson, G. B., & Burgess, C. (1985). Activation and Selection Processes in the Recognition of Ambiguous Words. *Journal of Experimental Psychology: Human Perception and Performance, 11*(1), 28-39.

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

Sliwinska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., & Devlin, J. T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Front Psychol, 3*, 161.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cerebral Cortex, 19*(7), 1493-1503.

Song, Y., Bu, Y., Hu, S., Luo, Y., & Liu, J. (2010). Short-term language experience shapes the plasticity of the visual word form area. *Brain research, 1316*, 83-91.

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2006). Converging language streams in the human temporal lobe. *J Neurosci, 26*(28), 7328-7336.

Stewart, L., Meyer, B., Frith, U., & Rothwell, J. (2001). Left posterior BA37 is involved in object recognition: a TMS study. *Neuropsychologia, 39*(1), 1-6.

Strand, F., Forssberg, H., Klingberg, T., & Norrelgen, F. (2008). Phonological working memory with auditory presentation of pseudo-words—an event related fMRI study. *Brain research, 1212*, 48-54.

Swaab, T., Brown, C., & Hagoort, P. (1998). Understanding ambiguous words in sentence contexts: electrophysiological evidence for delayed contextual selection in Broca's aphasia. *Neuropsychologia, 36*(8), 737-761.

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.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*(6), 645-659.

Swinney, D., Zurif, E., & Nicol, J. (1989). The Effects of Focal Brain Damage on Sentence Processing: an examination of the neurological organization of a mental module. *J Cogn Neurosci, 1*(1), 25-37.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language, 27*(3), 324-340.

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research, 49*(2-3), 161-167.

Tabossi, P., & Zardon, F. (1993). Processing Ambiguous Words in Context. *Journal of Memory and Language, 32*(3), 359-372.

Tahmasebi, A. M., Davis, M. H., Wild, C. J., Rodd, J. M., Hakyemez, H., Abolmaesumi, P., et al. (2012). Is the Link between Anatomical Structure and Function Equally Strong at All Cognitive Levels of Processing? *Cerebral Cortex, 22*(7), 1593-1603.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*(4), 427-440.

Terao, Y., Ugawa, Y., Suzuki, M., Sakai, K., Hanajima, R., Gemba-Shimizu, K., et al. (1997). Shortening of simple reaction time by peripheral electrical and submotor-threshold magnetic cortical stimulation. *Exp Brain Res, 115*(3), 541-545.

Thickbroom, G. W. (2007). Transcranial magnetic stimulation and synaptic plasticity: experimental framework and human models. *Experimental Brain Research, 180*(4), 583-593.

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 of the United States of America, 94*(26), 14792-14797.

Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D’Esposito, M., Kan, I. P., et al. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 109-120.

Thompson-Schill, S. L., Swick, D., Farah, M. J., D’Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proceedings of the National Academy of Sciences, 95*(26), 15855-15860.

Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia, 35*(10), 1319-1327.

Tranel, D., Kemmerer, D., Adolphs, R., Damasio, H., & Damasio, A. R. (2003). Neural correlates of conceptual knowledge for actions. *Cognitive Neuropsychology, 20*(3-6), 409-432.

Tremblay, P., Sato, M., & Small, S. L. (2012). TMS-induced modulation of action sentence priming in the ventral premotor cortex. *Neuropsychologia, 50*(2), 319-326.

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

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

Twomey, T., Kawabata Duncan, K. J., Price, C. J., & Devlin, J. T. (2011). Top-down modulation of ventral occipito-temporal responses during visual word recognition. *NeuroImage, 55*(3), 1242-1251.

Ueno, S., Tashiro, T., & Harada, K. (1988). Localized Stimulation of Neural Tissues in the Brain by Means of a Paired Configuration of Time-Varying Magnetic-Fields. *Journal of Applied Physics, 64*(10), 5862-5864.

Utman, J. A., Blumstein, S. E., & Sullivan, K. (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. *Brain and Language, 79*(3), 444-472.

van Turennout, M., Ellmore, T., & Martin, A. (2000). Long-lasting cortical plasticity in the object naming system. *Nature neuroscience, 3*(12), 1329-1334.

Vigneau, M., Beaucousin, V., Herve, P.-Y., Duffau, H., Crivello, F., Houde, O., et al. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage, 30*(4), 1414-1432.

Vu, H., Kellas, G., & Paul, S. T. (1998). Sources of sentence constraint on lexical ambiguity resolution. *Mem Cognit, 26*(5), 979-1001.

Vuong, L. C., & Martin, R. C. (2011). LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain and Language, 116*(1), 22-32.

Wagner, A. D., Paré-Blagoev, J. E., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31*(2), 329-338.

Wagner, T., Rushmore, J., Eden, U., & Valero-Cabre, A. (2009). Biophysical foundations underlying TMS: setting the stage for an effective use of neurostimulation in the cognitive neurosciences. *Cortex, 45*(9), 1025-1034.

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature reviews. Neuroscience, 1*(1), 73-79.

Walsh, V., Pascual-Leone, A., & Kosslyn, S. M. (2003). *Transcranial magnetic stimulation: a neurochronometrics of mind*: MIT press Cambridge, MA.

Wassermann, E. M., McShane, L. M., Hallett, M., & Cohen, L. G. (1992). Noninvasive mapping of muscle representations in human motor cortex. *Electroencephalogr Clin Neurophysiol, 85*(1), 1-8.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*(8), 989-994.

Wernicke, C. (1874). Der aphasische Symptomencomplex.

Wheatley, T., Weisberg, J., Beauchamp, M. S., & Martin, A. (2005). Automatic priming of semantically related words reduces activity in the fusiform gyrus. *Journal of Cognitive Neuroscience, 17*(12), 1871-1885.

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. *Cereb Cortex, 21*(4), 831-844.

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cereb Cortex, 21*(5), 1066-1075.

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cerebral Cortex, 17*(10), 2322-2333.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping, 4*(1), 58-73.

Xiang, H., Fonteijn, H. M., Norris, D. G., & Hagoort, P. (2010). Topographical functional connectivity pattern in the perisylvian language networks. *Cerebral Cortex, 20*(3), 549-560.

Xiao, Z., Zhang, J. X., Wang, X., Wu, R., Hu, X., Weng, X., et al. (2005). Differential activity in left inferior frontal gyrus for pseudowords and real words: An event-related fMRI study on auditory lexical decision. *Human brain mapping, 25*(2), 212-221.

Yee, E., Blumstein, S. E., & Sedivy, J. C. (2008). Lexical-Semantic Activation in Broca's and Wernicke's Aphasia: Evidence from Eye Movements. *Journal of Cognitive Neuroscience, 20*(4), 592-612.

Yoncheva, Y. N., Zevin, J. D., Maurer, U., & McCandliss, B. D. (2010). Auditory selective attention to speech modulates activity in the visual word form area. *Cerebral Cortex, 20*(3), 622-632.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*(5058), 846-849.

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

Zhu, Z., Hagoort, P., Zhang, J. X., Feng, G., Chen, H., Bastiaansen, M., et al. (2012). The anterior left inferior frontal gyrus contributes to semantic unification. *NeuroImage, 60*(4), 2230-2237.

Appendix

Experimental Sentences and Probe Words.

In the ambiguous sentences, the ambiguous word is underlined and the disambiguating information is italicised. A = Ambiguous, U = Unambiguous, Rel. = Related, Unrel. = Unrelated, Inapp. = Inappropriate.

			Probe	
	Sentence	**Rel.**	**Unrel.**	**Inapp.**
A	the students believed that the plant should never have been *built*	manufacture	launch	bush
U	the students believed that the hill could never have been climbed	valley	dollar	
A	the couple hoped that the court they had chosen was good for *tennis*	arena	backward	lawyer
U	the couple hoped that the church they had chosen was good for weddings	chapel	hazard	
A	the man chose the new foil and hoped that it would improve his *fencing*	sword	baboon	oven
U	the man chose the new herbs and hoped that they would improve his cooking	parsley	clutter	
A	when the mother saw the clip, she knew that her daughter would like the *movie*	scene	rum	clasp
U	when the mother finished the scarf, she hoped that her daughter would like the colour	wool	fork	

A	the woman explained that although the ticks were small they were particularly *harmful*	mite	temple	correct
U	the woman explained that although the stains were small they were particularly stubborn	dirty	future	
A	her boyfriend explained that the deed he had seen had been extremely *courageous*	action	table	contract
U	her boyfriend explained that the pond he had seen had been extremely *overgrown*	nature	delay	
A	the man found the file difficult to use as he had never worked with *metal*	tool	belly	folder
U	the man found the lake difficult to find as he had never used a compass	lagoon	sleek	
A	her husband found the coat difficult to put on with his new *paint-roller*	layer	history	jacket
U	her husband found the bus difficult to count on with his new shift-pattern	transport	wisdom	
A	the woman had to make the toast with a very old *microphone*	speech	blanket	bread
U	the student had to wrap the wrist with a very old bandage	arm	rock	
A	the company was certain that the post would be quickly *filled*	employ	slender	package
U	the family were certain that the farm would be quickly sold	ranch	sunset	

A	she was worried that the coach they had chosen was too *inexperienced*	instructor	lemon	vehicle
U	she was worried that the path they had chosen was too inaccessible	route	funny	
A	as soon as the woman noticed the large mole, she called a local *doctor*	blemish	carnival	agent
U	as soon as the woman noticed the broken fridge, she called a local plumber	appliance	sincerity	
A	the man believed that the scoop given to his daughter would make the *news*	gossip	picnic	shovel
U	the man thought that the blouse bought for his girlfriend would suit her taste	clothing	engineer	
A	his girlfriend was surprised that the punch was not at all *fruity*	drink	briefcase	beat
U	her boyfriend was surprised that the mouse was not at all scary	rat	strategy	
A	The woman was surprised to find a passage that had not yet been *translated*	text	badminton	corridor
U	The woman was surprised to find a request that had not yet been processed	claim	earth	
A	the man was surprised that the speaker was not going to be *repaired*	amplify	olive	lecture
U	the man was surprised that the apple was not going to be washed	snack	switch	

A	The boy picked up the straw and put it in his father's *barn*	hay	close	tube
U	The boy picked up the frog and put it in his sister's hands	amphibian	assorted	
A	the student lost the cap that went with her favourite *biro*	lid	kettle	headgear
U	the student visited her aunt who lived in her favourite city	relative	audience	
A	the housewife knew that the dates she had chosen would be *delicious*	food	pilot	diary
U	the housewife knew that the wounds she had suffered would be painful	injury	city	
A	the man explained that the pane was quite old and was from an earlier *window*	glass	queen	hurt
U	the man explained that the team was quite new and was from a different league	player	sketch	
A	the woman explained that the plot was not expected to be used for *housing*	land	sip	narrative
U	the woman explained that the cake was not intended to be eaten for lunch	pudding	cubicle	

A	The man was surprised that the panel that they had seen was *wooden*	plank	waste	committee
U	The man was annoyed that the hammer that he had bought was broken	mallet	ten	
A	the man was told that the compound contained many separate *apartments*	estate	guilt	chemistry
U	The boy was told that the ocean contained many different animals	sea	power	
A	although he had bought the bulbs, he was not sure how they should be *planted*	flower	chat	electric
U	although he had noticed the ants, he was not sure how they could be removed	insect	poem	
A	their grandfather explained that the corn had always been very *painful*	wart	magazine	cereal
U	his grandchildren thought that the mask had always been very scary	disguise	treacle	
A	the businessman told him that the fan had been very *supportive*	admirer	search	cool
U	the businessman told him that the coal had been very expensive	heating	surgeon	
A	his grandmother thought that the boxer was too vicious for a *pet*	dog	link	sport
U	his grandmother thought that the brandy was too expensive for a gift	liqueur	interview	

A	he believed that the pack contained between five and ten *wolves*	herd	cheap	card
U	he noticed that the bone had between one and three fractures	limb	whistle	
A	the man was not sure if the deck he had chosen had been *shuffled*	casino	golden	floor
U	the man was not sure if the goat he was feeding had been groomed	livestock	ambulance	
A	the teacher explained that the knight in the story was very *chivalrous*	armour	rectangle	darkness
U	the teacher explained that the girl in the book was very stubborn	youngster	crucial	
A	his friend argued that the boy's cue was far too short for *snooker*	stick	library	prompt
U	his friend argued that the boy's tent was far too flimsy for camping	shelter	reverse	
A	the teacher explained that the bark was going to be very *damp*	wood	pound	sound
U	the teacher explained that the steam was going to be very hot	vapour	token	
A	the woman asked her husband to choose the pear that was really *juicy*	fruit	bronze	set
U	the teacher asked her students to choose a task that was fairly easy	activity	double	

A	the workers were told that every nail needed to be clean and *manicured*	finger	doubt	iron
U	the students were told that the song needed to be long and melodic	tune	wage	
A	the student wanted to find a cymbal to add to his *drumkit*	metal	gentle	sign
U	the student wanted to find a feather to add to the hat	accessory	detective	
A	the woman could not find a suitable spade among her collection of *cards*	black	petticoat	dig
U	the woman could not find a suitable shrub among her collection of plants	greenery	nostalgia	
A	the woman explained that the figure had been made from *porcelain*	ornament	bacon	graph
U	the woman explained that the coffee had been made with cream	caffeine	parade	
A	the student knew which flour should be used for that type of *baking*	ingredient	squeeze	petal
U	the student knew which weapon should be used for that type of fighting	pistol	flamenco	

A	everyone believed that the lobby was going to become very *powerful*	petition	salmon	hallway
U	everyone believed that the cattle was going to become very aggressive	bull	birthday	
A	the woman declared that noone should discuss the race of the future *applicants*	ethnic	mushroom	athlete
U	the woman declared that noone should contact the chief of their local department	manager	massive	
A	the man was told that an organ was not available for the *choir*	piano	novel	body
U	the man was told that the prayer was not suitable for the children	religion	employer	
A	the husband explained why the new ruler was needed by the *country*	emperor	dance	straight
U	the husband explained that the new leaflet was needed for the business	brochure	myth	
A	the newspaper article about the star was interesting for all *astronomers*	solar	second	famous
U	the newspaper article about the milk was interesting for all consumers	dairy	mile	
A	the student loved the model from the first moment he saw *her*	fashion	tractor	version
U	the woman loved the forest from the first moment she saw it	woodland	crisis	

A	the man was told that the chips were not suitable for his *computer*	electronic	delicate	potato
U	the girl was told that the nests were not unusual for that species	bird	epidemic	
A	the small child was certain that his chest was full of *treasure*	storage	journey	cardiac
U	the small child was certain that the tribe was full of warriors	community	knuckle	
A	the old woman thought that the sentence would be considered *ungrammatical*	word	antifreeze	justice
U	the old man thought that the writer would be considered literary	book	whole	
A	the old woman noticed that the bonnet was covered with *embroidery*	hat	beard	engine
U	the old woman noticed that the cactus was covered with pollen	prickly	laughter	
A	she was pleased to discover that the mould was suitable for making *statues*	shape	circle	fungus
U	she was pleased to discover that the soap was suitable for washing crockery	detergent	wealth	
A	her neighbour was worried that the spring might start to become *uncoiled*	elastic	nutrient	waterfall
U	her neighbour was worried that the burn might start to become infected	blister	giraffe	

A	the teacher explained that the clause could not easily be *translated*	phrase	greedy	sharp
U	the teacher explained that the snake could not easily be tamed	reptile	horizon	
A	the schoolgirl was told that the poker was a very important *heirloom*	fireplace	vinegar	betting
U	the schoolgirl was told that the rabbit was a very timid animal	bunny	scheme	
A	the teacher said that the root was surprisingly long for a *tree*	soil	culture	direction
U	the teacher knew that the guess was unusually good for a child	estimate	goldfish	
A	her elderly grandmother enjoyed the peal that came from the *cathedral*	bell	damage	rind
U	her elderly grandmother described the grief that came from her bereavement	sorrow	chemical	
A	the man was surprised that the cricket produced so much chirping	locust	laundromat	batsman
U	the man was surprised that the meadow produced so many butterflies	pasture	hostess	
A	the man thought that the club he had chosen was best for that *shot*	equipment	mystery	society
U	the man thought that the seat he had chosen was best for that concert	chair	north	

A	the professor did not know why the match could not be *lit*	strike	agree	game
U	the professor did not know why his lunch could not be found	eat	perfect	
A	the woman had read about the mousse in the new *cookbook*	dessert	kindness	deer
U	the woman had heard about the elk in the local zoo	animal	mistake	
A	the young woman noticed that the palm had several large *fruits*	tropical	public	fist
U	the young woman noticed that her chin had several large spots	face	home	
A	the young girl thought that the band was just perfect for the *hat*	cloth	tremble	musician
U	the young girl thought that the gift was just perfect for her friend	present	morning	
A	the men thought that the port would not be suitable for the *dinner*	sherry	spa	cargo
U	the men thought that the trout would not be suitable for the dinner	fish	jumper	
A	the woman said that the jam was on the other *motorway*	traffic	bedroom	preserve
U	the woman said that the soup was on the other menu	liquid	arson	
A	the man was annoyed that the horn had been damaged by the *mechanic*	beep	towel	antler
U	the man was surprised that the nut had been cracked by the squirrel	seed	beer	

A	the man was surprised that the log contained so much *information*	journal	hen	branch
U	the man was surprised that the cage contained so many animals	confine	keyboard	
A	the children were told that the bat they had seen was a *carnivore*	wings	stale	ball
U	the children were told that the fox they had seen was a herbivore	vixen	main	
A	the man explained that the pen was mainly used for the *livestock*	enclosure	migraine	ballpoint
U	the man explained that the hut was mainly used by the climbers	cabin	thumb	
A	their mother told them that the tale was quite long but very *exciting*	story	pavement	rear
U	their mother told them that the meat was quite tough but very tasty	beef	duty	
A	the woman was not sure whether the stork was strong enough to *fly*	heron	baker	stem
U	the woman was not sure whether the glove was clean enough to wear	mitten	goblet	
A	the man was worried that the bale was too large for the *trailer*	harvest	silver	prisoner
U	the man was worried that the jug was too small for the cocktail	pour	owl	
A	The teenager looked at the litter and decided to take one *kitten*	offspring	commerce	rubbish
U	The teenager went to the wedding and decided to take many photos	ceremony	remain	

A	the man would need to replace the pipe or he would have to stop *smoking*	mouth	radio	water
U	the man would need to avoid the storm or he would have to change clothes	rain	wine	
A	the man knew that one more ace would be enough to win the game of *tennis*	score	autumn	diamond
U	the man thought that one more bet could be enough to win the prize of money	gamble	noon	
A	the girl understood that her genes had all come from her *ancestors*	biology	gutter	denim
U	the girl believed that her brains had all come from her mother	clever	custard	
A	the woman told her daughter that currants were usually very *tasty*	raisin	pin	tide
U	the woman told her daughter that winters were usually very cold	season	fuss	
A	according to the teacher, the prophet was mentioned in the new *testament*	preach	blue	finance
U	according to the teacher, the colonel was mentioned in the local newspaper	military	six	
A	the man explained that the brake was not good enough for his *lorry*	mechanic	lady	rest
U	the man believed that his wife was not kind enough to the children	marriage	passport	

A	the man explained why the volume was not going to be *published*	edition	flannel	loudness
U	the man explained why the kingdom was not going to be conquered	empire	sponge	
A	the woman was worried when she saw the steak, as it had been *undercooked*	sirloin	cassette	wager
U	the woman was worried when she saw the bench, as it had been vandalised	seating	colon	
A	the man really wanted mussels but he didn't know how they should be *cooked*	shellfish	west	fitness
U	the man really hated battles as he didn't know how they could be won	conflict	distance	
A	the man was told that the cast would be chosen by an experienced *doctor*	plaster	ivy	actor
U	the man was told that the bath would be chosen by an experienced designer	tub	museum	
A	the man worried that the chord would not fit the *melody*	note	economic	rope
U	the man hoped that the scar would not affect his confidence	mark	left	
A	the woman had just noticed the flair of the girl's *writing*	talent	unit	trousers
U	the woman had just noticed the strap of the girl's rucksack	handle	hope	

A	she was not sure whether the fir could be used for *timber*	tree	puppet	mink
U	she was not sure whether the skirt could be used for school	garment	honest	
A	the man did not know how the interest had always been *calculated*	money	sickness	attention
U	the man did not know how the worker had always been tolerated	employee	swing	
A	the teacher told the young girl that the hare was not a *pet*	wildlife	forecast	wig
U	the teacher told the young girl that the town was not a city	urban	peach	
A	the woman was told that the mint was used for making *coins*	factory	gym	flavour
U	the woman was told that the crate was used for storing goods	container	napkin	
A	the young woman knew that her pupils had started to *dilate*	eye	talk	student
U	the young woman knew that the mirror had started to crack	reflect	ticket	
A	the woman was annoyed that the mail was not always *delivered*	parcel	universe	masculine
U	the woman was annoyed that the shop was not always advertised	retail	oil	
A	the young man was not sure if the bass would need to be *tuned*	guitar	news	bottom
U	the young man was not sure if the tooth would need to be drilled	dental	addiction	

A	the expert explained that the bolt had come from the base of a *thunderstorm*	lightning	draw	secure
U	the detective noticed that the spy had walked to the front of the aeroplane	espionage	overflow	